Grid technologies for large-scale projects

N. S. Astakhov, A. S. Baginyan, S. D. Belov, A. G. Dolbilov, A. O. Golunov, I. N. Gorbunov, N. I. Gromova, I. A. Kashunin, V. V. Korenkov, V. V. Mitsyn, V. V. Palichik, S. V. Shmatov, T. A. Strizh, E. A. Tikhonenko, V. V. Trofimov, N. N. Voitishin, V. E. Zhiltsov

Laboratory of Information Technologies

JINR

Cluj-Napoca

2015

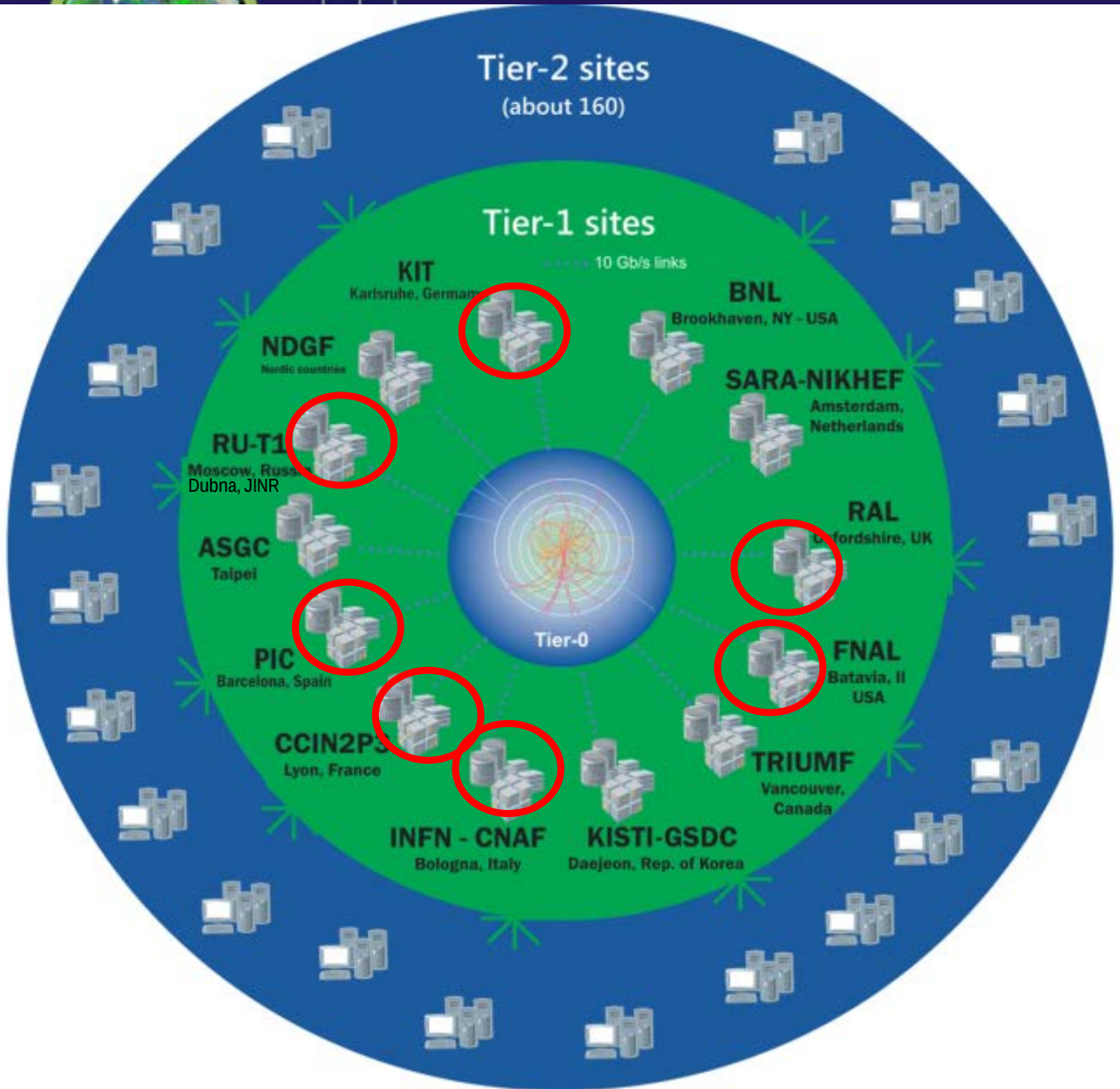
Grid technologies - a way to success

On a festivity dedicated to receiving the Nobel Prize for discovery of Higgs boson, CERN Director professor Rolf Dieter Heuer directly called the **grid-technologies one of three pillars of success** (alongside with the LHC accelerator and physical installations).

Without implementation of the grid-infrastructure on LHC it would be impossible to process and store enormous data coming from the collider and therefore to make discoveries.

Nowadays, every large-scale project will fail without using a distributed infrastructure for data processing.

LHC Computing Model



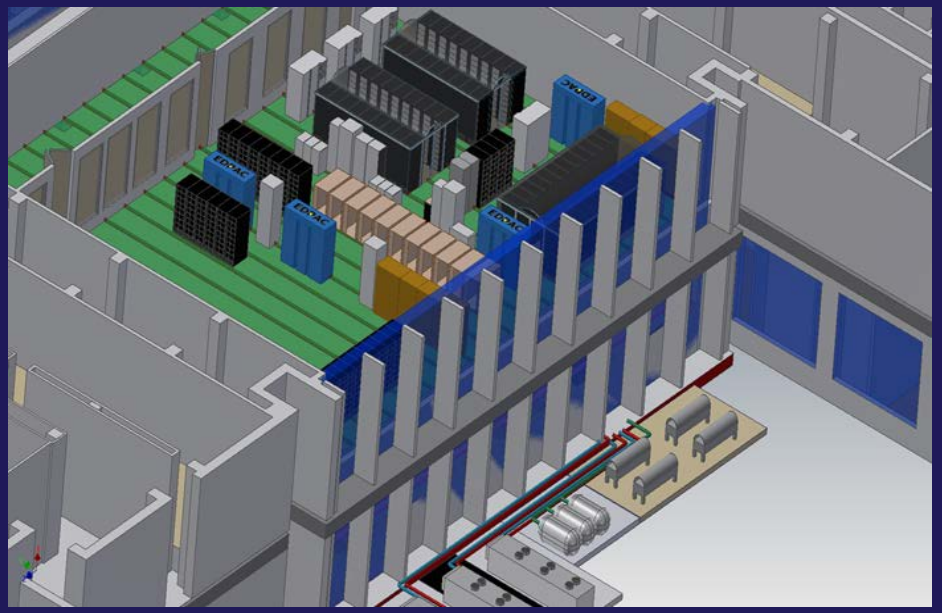
Tier-0 (CERN):
- Data recording
- Initial data reconstruction
- Data distribution

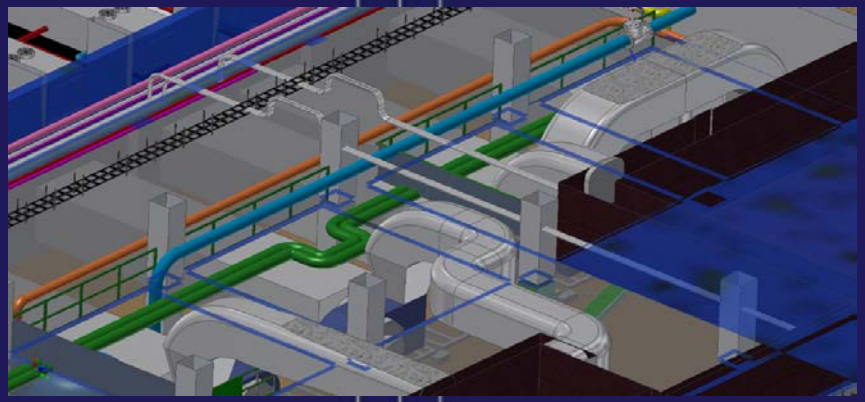
Tier-1 (11→14 centres):
- Permanent storage
- Re-processing
- Analysis
- Simulation

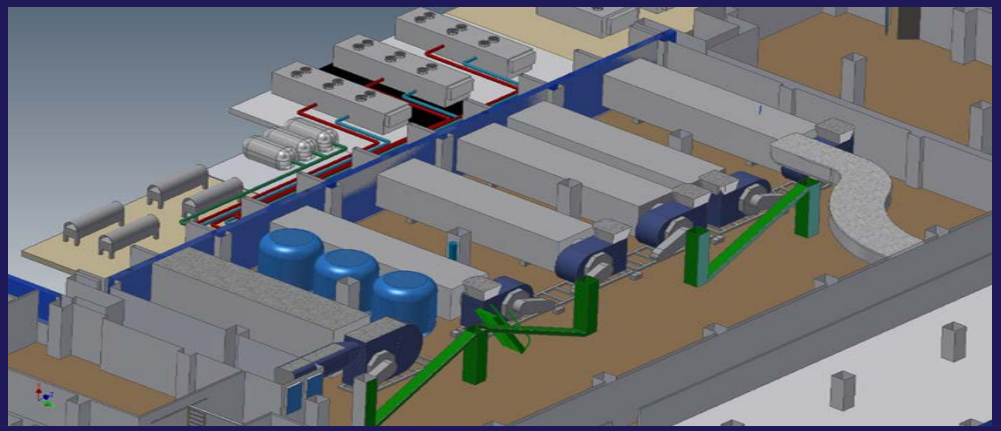
Tier-2 (>200 centres):
- Simulation
- End-user analysis

Creation of CMS Tier1 in JINR

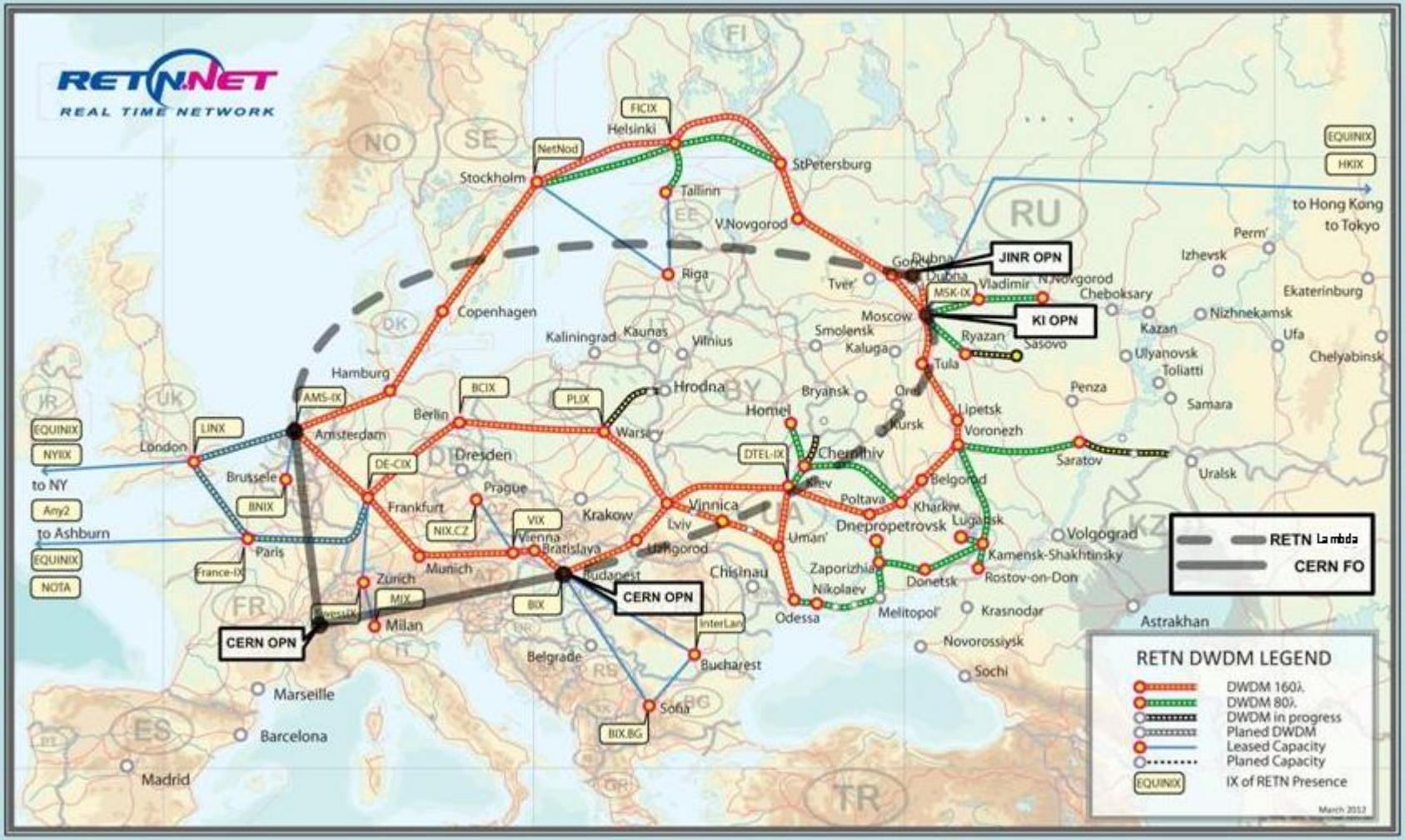
- Engineering infrastructure (a system of uninterrupted power supply, climate - control);
- High-speed reliable network infrastructure with a dedicated reserved data link to CERN (LHCOPN);
- Computing system and storage system on the basis of disk arrays and tape libraries of high capacity;
- 100% reliability and availability.



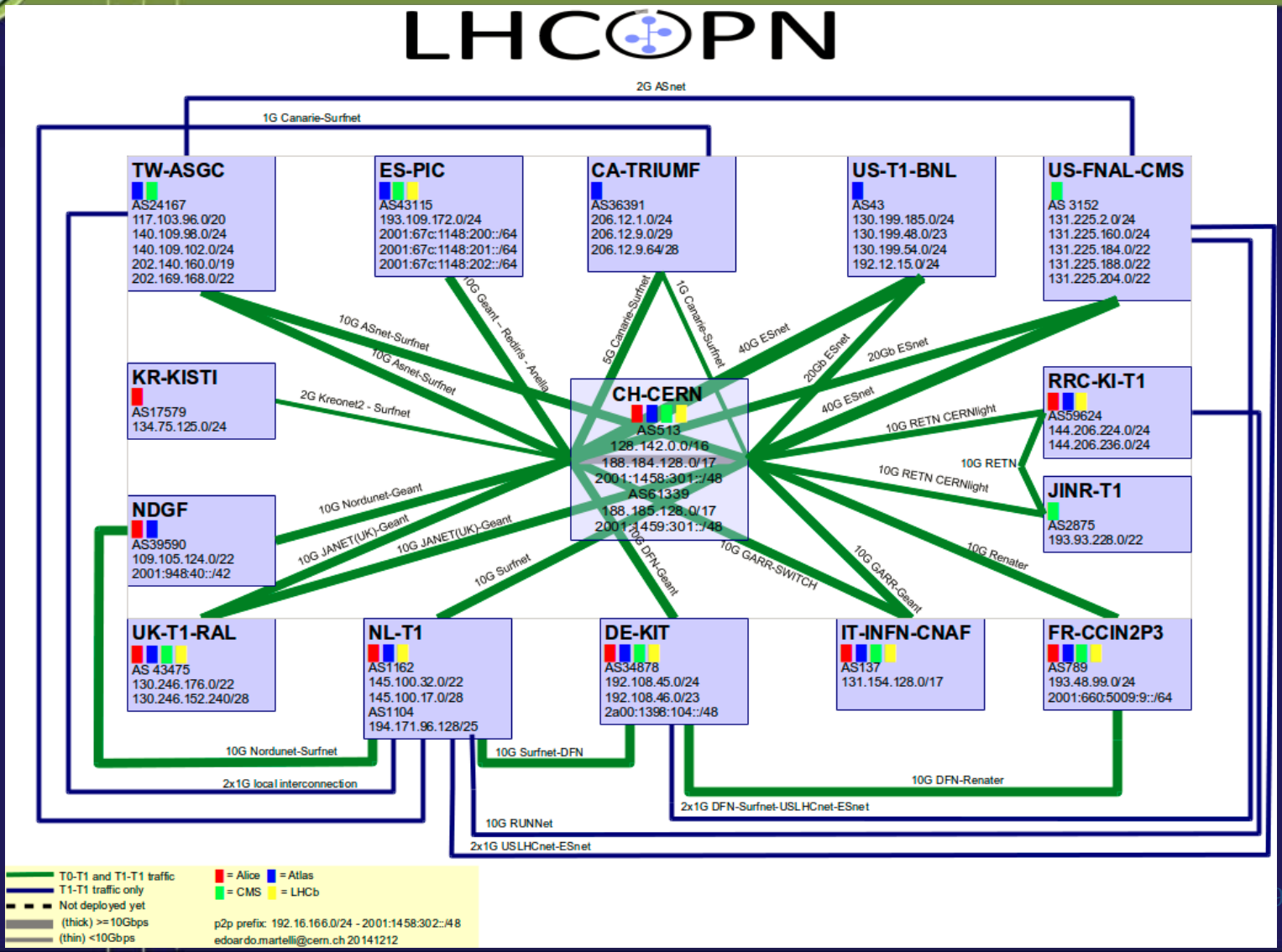




JINR Tier-1 Connectivity Scheme



LHCOPN - The Optical Private Network for the Large Hadron Collider

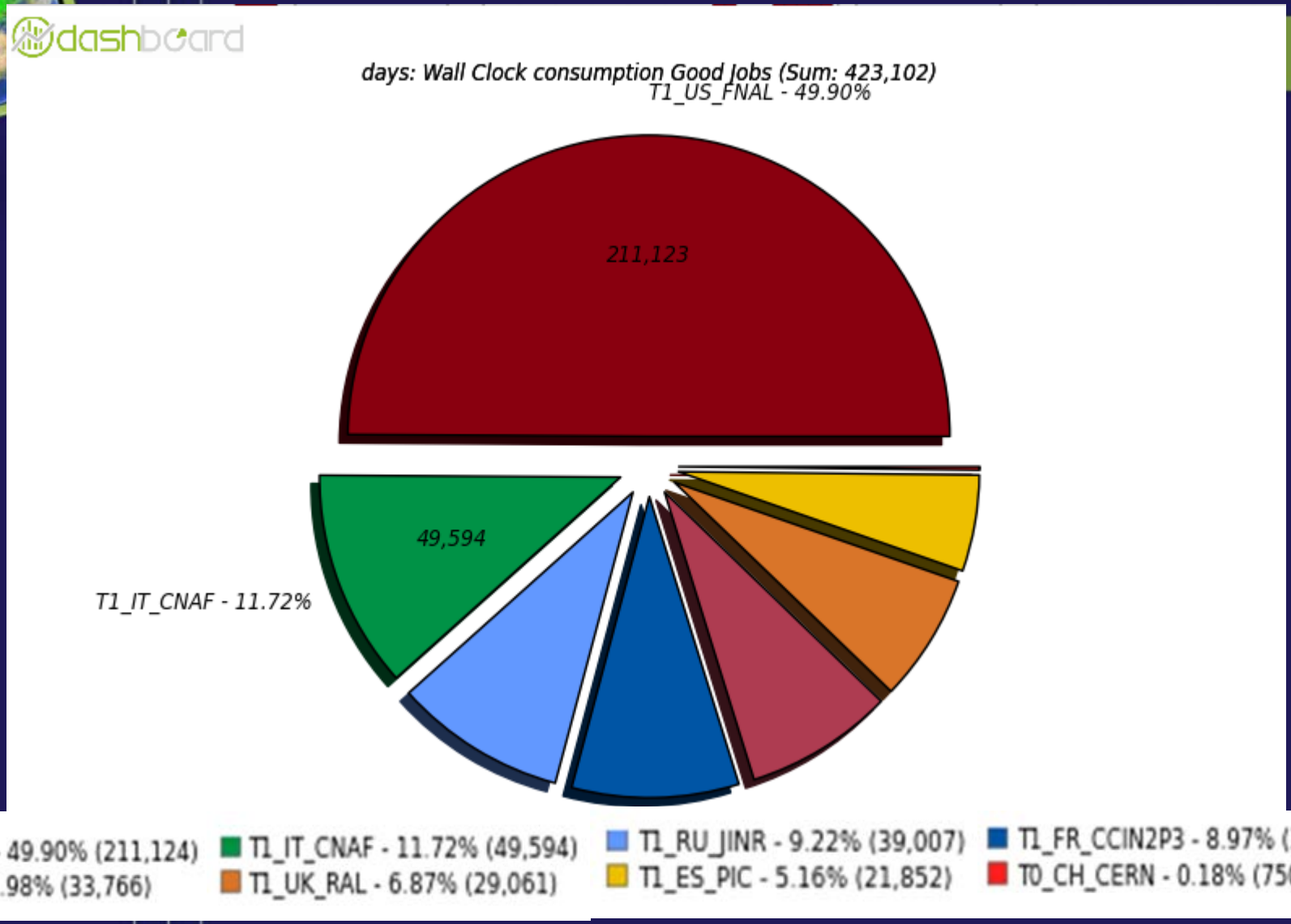


Components of the Tier1

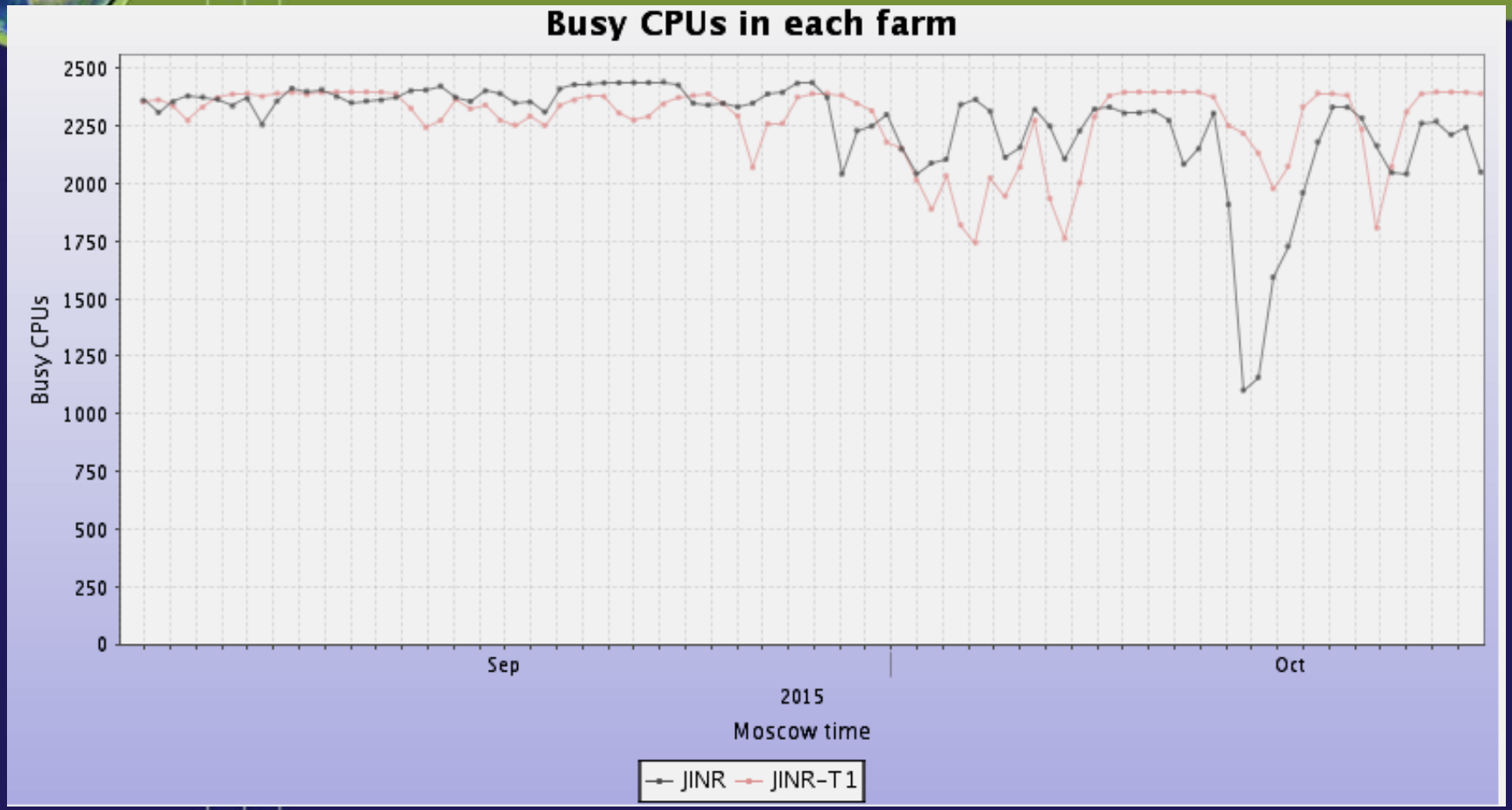
Current configuration and plans


Russia:
NRC KI
US-BNL
Amsterdam/NIKHEF-SARA
Bologna/CNAF
Taipei/ASGC
Ca-
TRIUMF
JINR
NDGF
CERN
US-FNAL
Barcelona/PIC
Lyon/CCIN2P3
UK-RAL
De-FZK

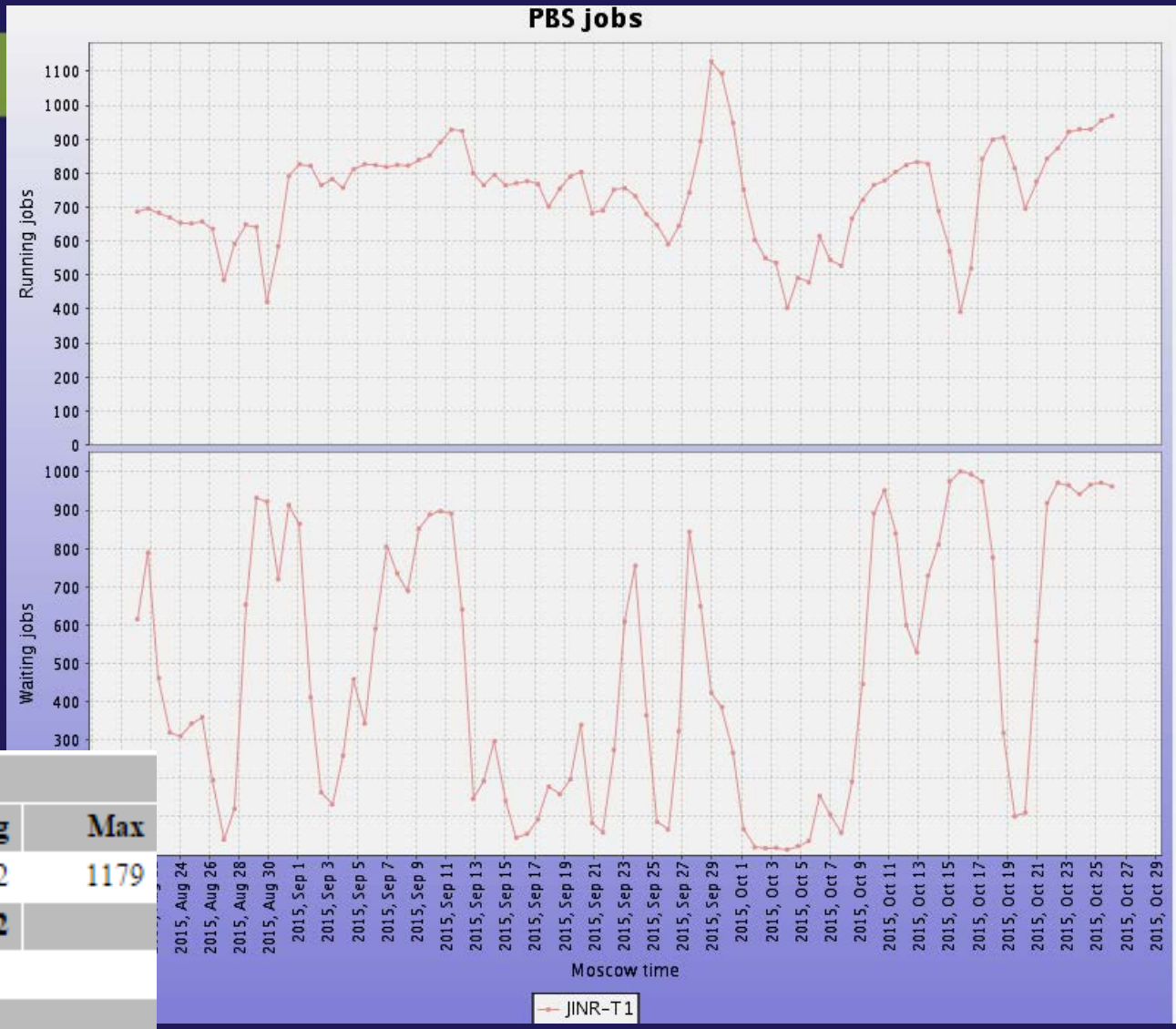
Last month normalized CPU time by CMS Tier-1



JINR GRID farm usage history



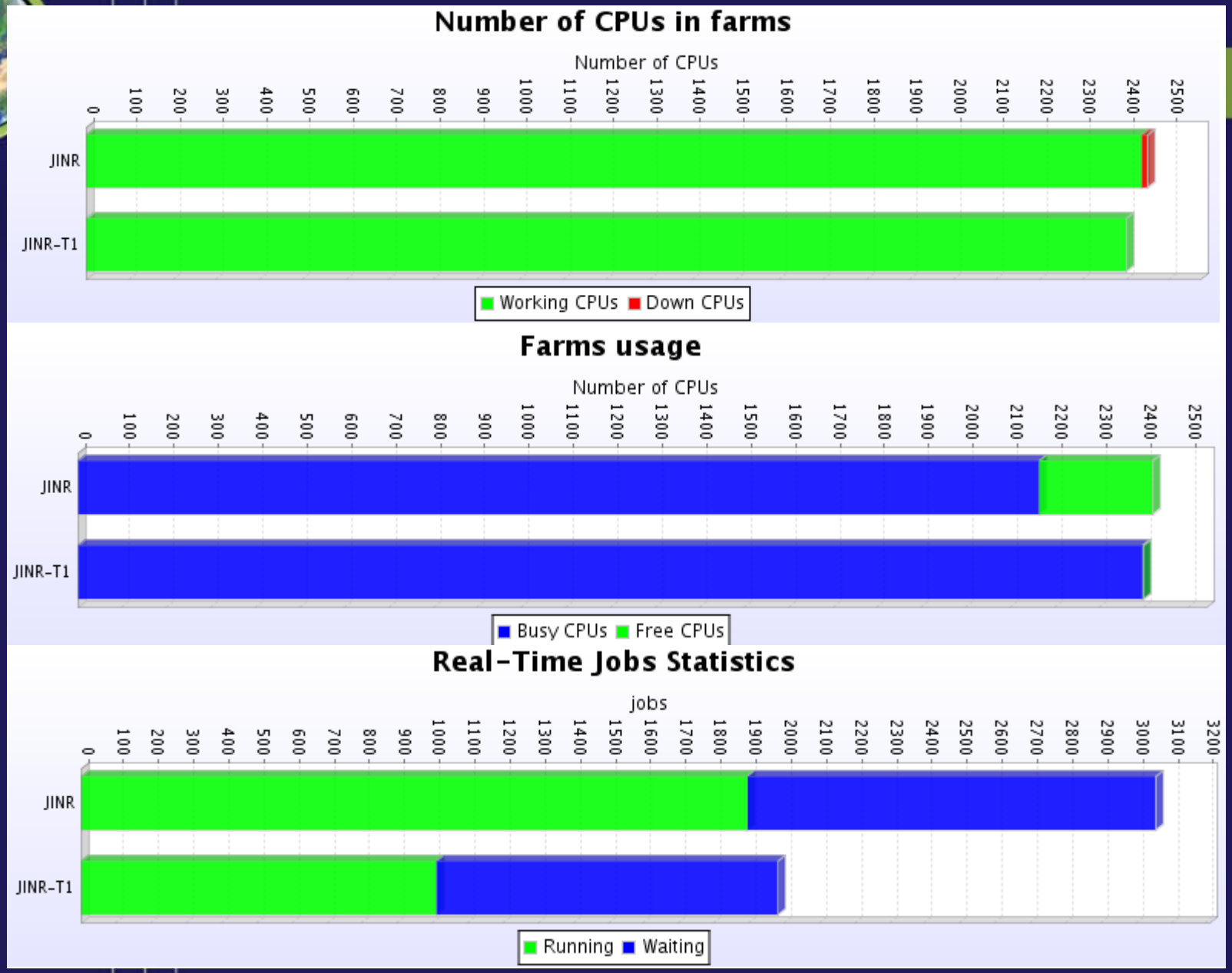
Tier-1 jobs (history Sept-Oct 2015)



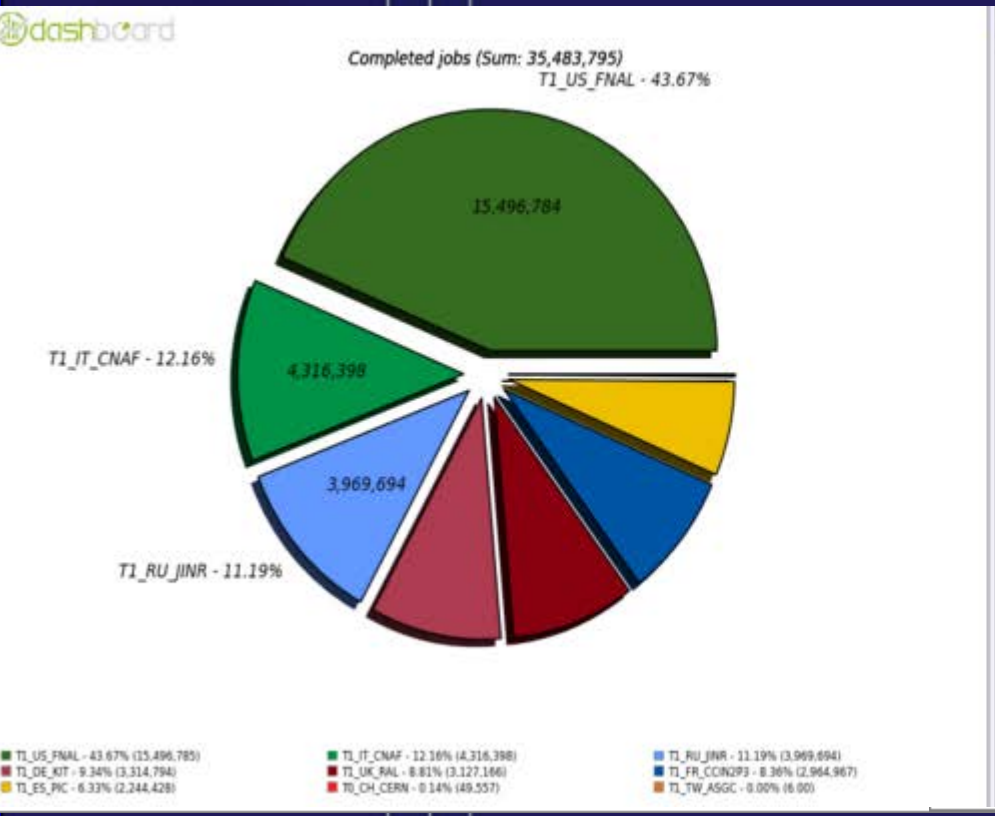
Running jobs				
Farm	Last value	Min	Avg	Max
JINR-T1	969	0	742	1179
Total	969		742	

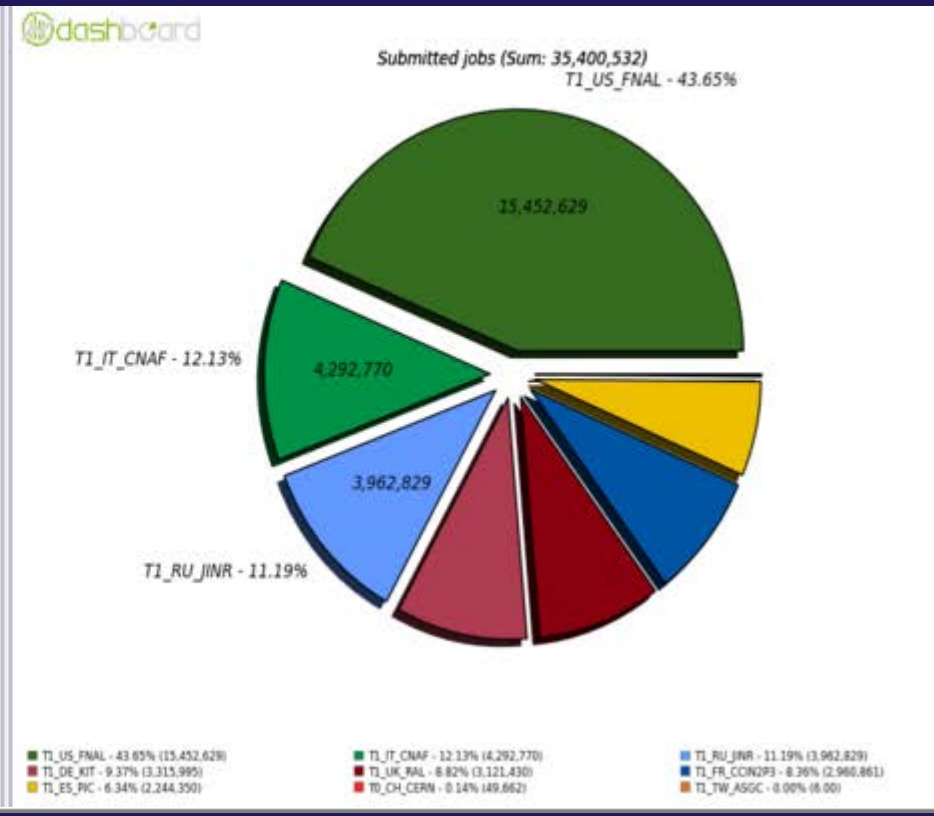
Waiting jobs				
Farm	Last value	Min	Avg	Max
JINR-T1	962	0	476.3	1054

Tier-1 & Tier-2 usage and jobs (real time)

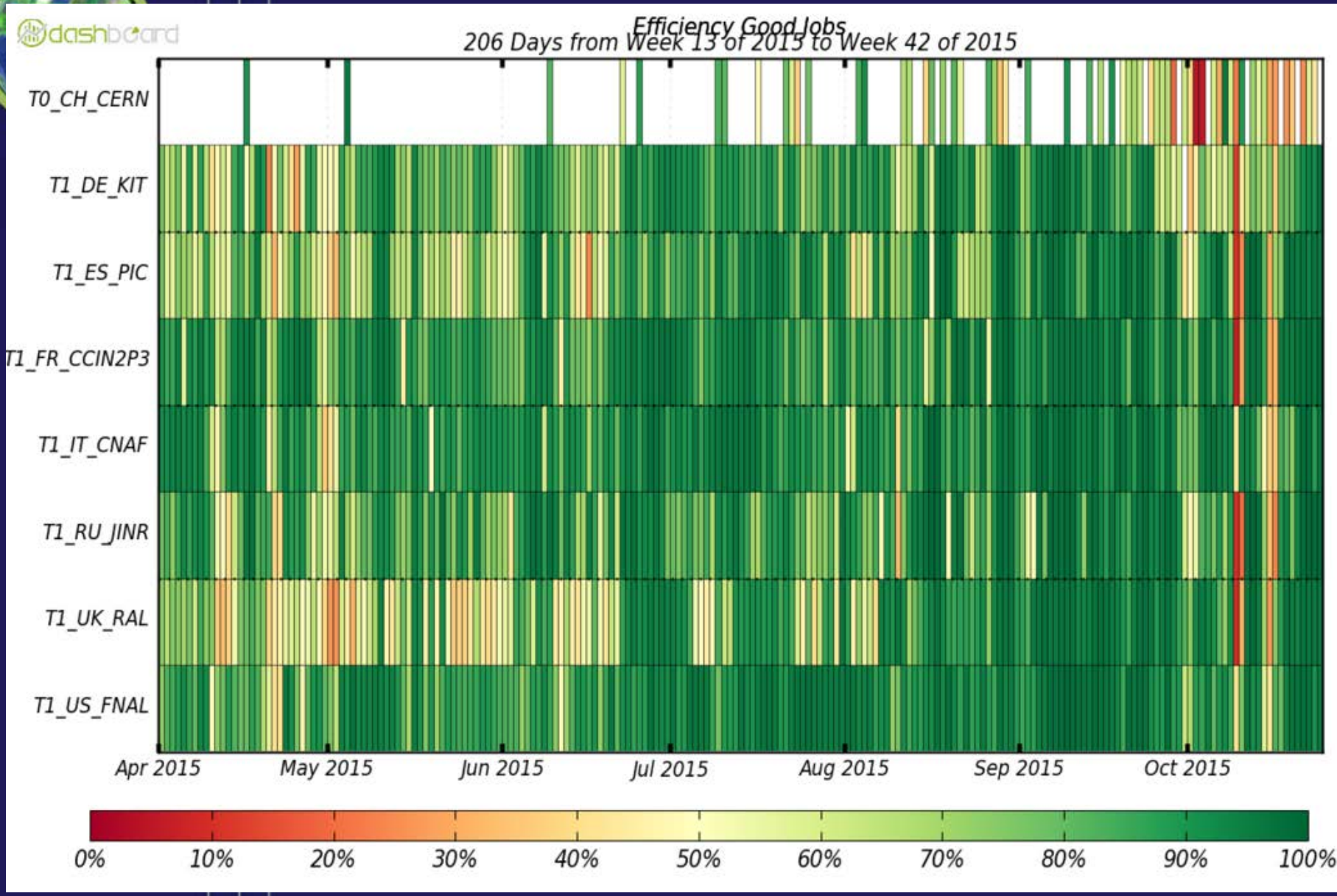


March – October 2015 jobs

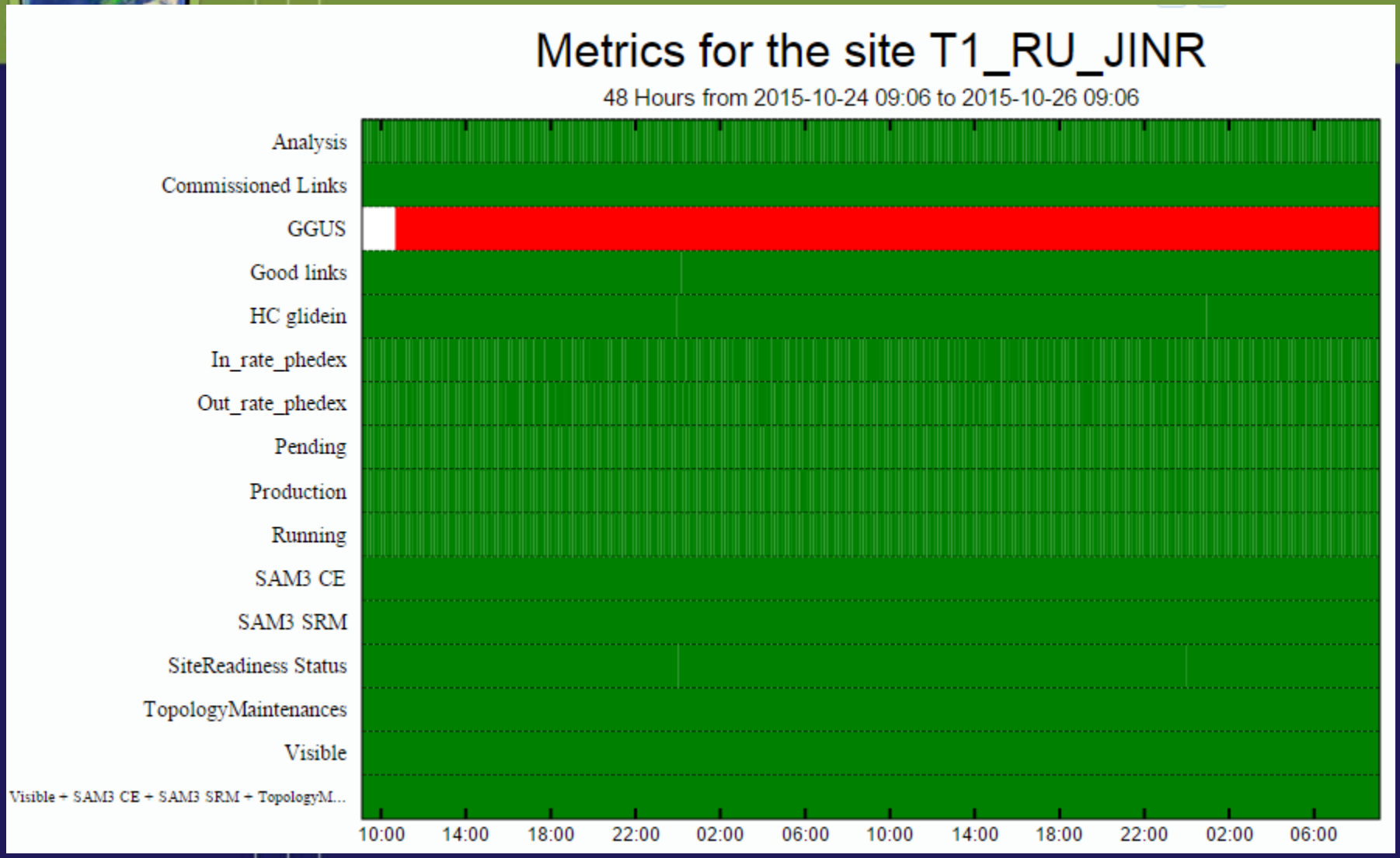




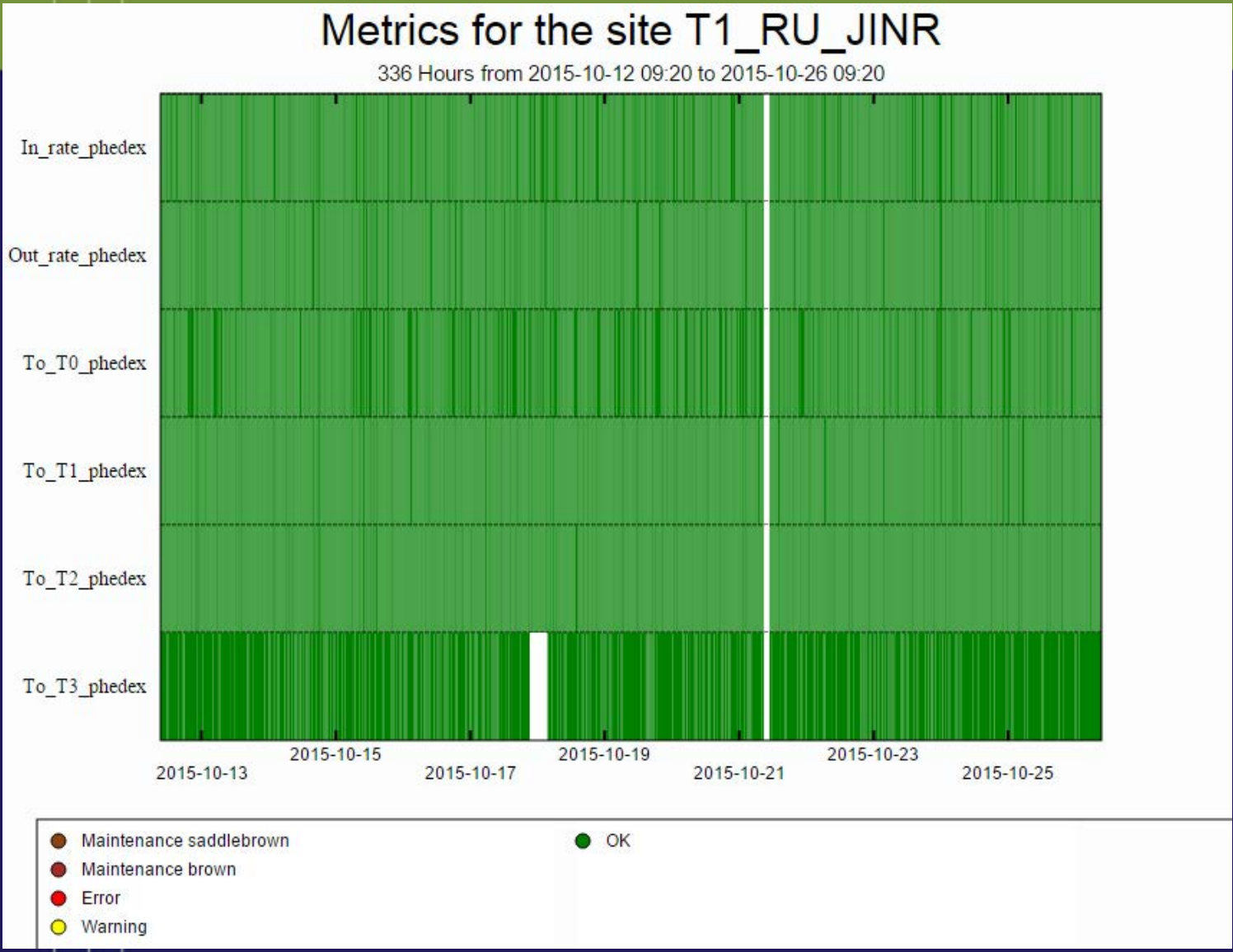
Job Efficiency



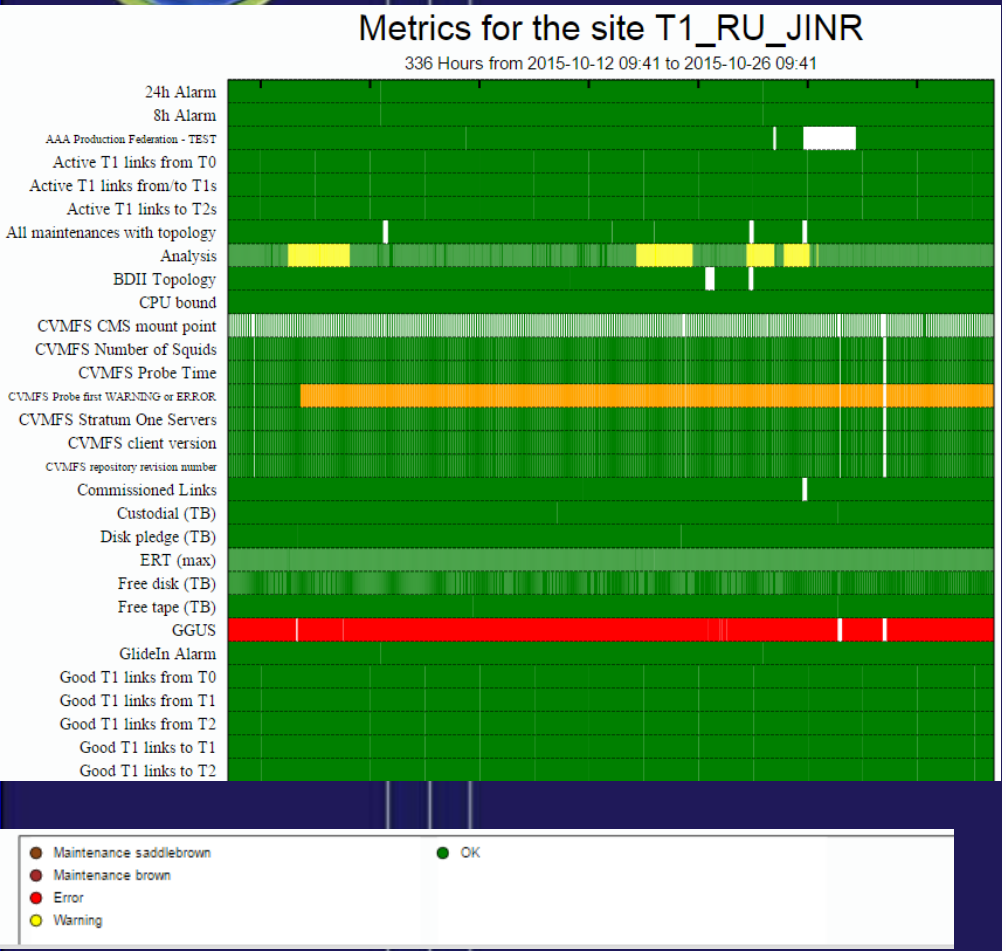
Default metrics

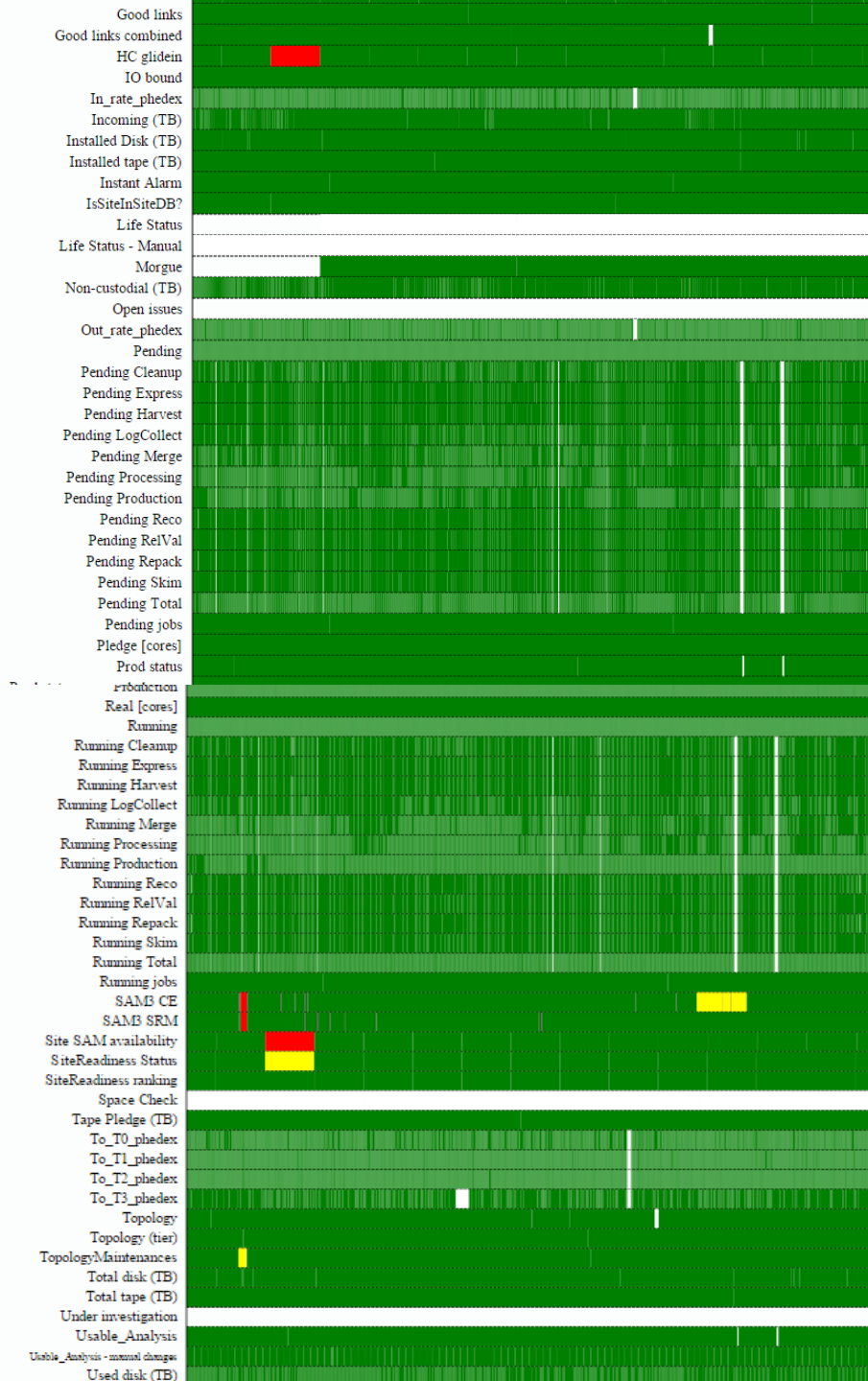


TRANSFERS


Metrics for the site T1_RU_JINR
336 Hours from 2015-10-12 09:20 to 2015-10-26 09:20
In_rate_phedex
Out_rate_phedex
To_T0_phedex
To_T1_phedex
To_T2_phedex
To_T3_phedex
2015-10-13
2015-10-15
2015-10-17
2015-10-19
2015-10-21
2015-10-23
2015-10-25
Maintenance saddlebrown
Maintenance brown
Error
Warning
OK

ALL METRICS


Metrics for the site T1_RU_JINR
336 Hours from 2015-10-12 09:41 to 2015-10-26 09:41
24h Alarm
8h Alarm
AAA Production Federation - TEST
Active T1 links from T0
Active T1 links from/to T1s
Active T1 links to T2s
All maintenances with topology
Analysis
BDII Topology
CPU bound
CVMFS CMS mount point
CVMFS Number of Squids
CVMFS Probe Time
CVMFS Probe first WARNING or ERROR
CVMFS Stratum One Servers
CVMFS client version
CVMFS repository revision number
Commissioned Links
Custodial (TB)
Disk pledge (TB)
ERT (max)
Free disk (TB)
Free tape (TB)
GGUS
GlideIn Alarm
Good T1 links from T0
Good T1 links from T1
Good T1 links from T2
Good T1 links to T1
Good T1 links to T2
Maintenance saddlebrown
Maintenance brown
Error
Warning
OK


Good links
Good links combined
HC glidein
IO bound
In_rate_phedex
Incoming (TB)
Installed Disk (TB)
Installed tape (TB)
Instant Alarm
IsSiteInSiteDB?
Life Status
Life Status - Manual
Morgue
Non-custodial (TB)
Open issues
Out_rate_phedex
Pending
Pending Cleanup
Pending Express
Pending Harvest
Pending LogCollect
Pending Merge
Pending Processing
Pending Production
Pending Reco
Pending RelVal
Pending Repack
Pending Skim
Pending Total
Pending jobs
Pledge [cores]
Prod status
Real [cores]
Running
Running Cleanup
Running Express
Running Harvest
Running LogCollect
Running Merge
Running Processing
Running Production
Running Reco
Running RelVal
Running Repack
Running Skim
Running Total
Running jobs
SAM3 CE
SAM3 SRM
Site SAM availability
SiteReadiness Status
SiteReadiness ranking
Space Check
Tape Pledge (TB)
To_T0_phedex
To_T1_phedex
To_T2_phedex
To_T3_phedex
Topology
Topology (tier)
TopologyMaintenances
Total disk (TB)
Total tape (TB)
Under investigation
Usable_Analysis
Usable_Analysis - manual changes
Used disk (TB)

Site Readiness Status

T1_RU_JINR

Site Readiness Status:	R	R	R	W	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Daily Metric:	O	O	O	O	O	O	O	O	O	E	O	O	O	O	O	O	O	O	O	O	O	O
Maintenance:	Up	Up	Up	Up	Up	Up	Up	Up	Up	~	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up
HammerCloud:	100%	100%	100%	100%	100%	100%	100%	100%	100%	76%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
SAM Availability:	100%	92%	100%	100%	100%	100%	100%	100%	100%	87%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Good T1 links from T0:	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	4/4	4/4	4/4	4/4	
Good T1 links from T1s:	23/23	23/23	23/23	20/23	23/23	24/24	24/24	23/23	23/23	24/24	23/23	23/23	23/23	23/23	23/23	23/23	23/23	25/25	26/26	26/26	26/26	
Good T1 links from T2s:	39/49	39/50	59/75	80/94	76/94	78/91	77/92	77/92	76/92	77/94	76/92	75/89	74/90	78/92	74/90	78/94	75/94	72/94	69/93	75/90	75/92	
Good T1 links to T1s:	28/28	27/28	26/28	25/28	28/28	28/28	28/28	28/28	27/28	28/28	28/28	28/28	28/28	28/28	28/28	28/28	28/28	30/30	30/30	30/30	30/30	
Good T1 links to T2s:	37/46	39/49	40/47	41/48	42/51	42/50	38/48	39/50	42/50	42/49	41/49	43/50	41/50	45/52	43/52	49/60	49/59	71/87	66/87	73/85	78/87	
Active T1 links from T0:	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	4	4	
Active T1 links from/to T1s:	16(d)-13(u)	16(d)-13(u)	16(d)-13(u)	16(d)-13(u)	16(d)-13(u)	16(d)-13(u)	16(d)-13(u)	16(d)-13(u)	18(d)-13(u)	20(d)-13(u)	24(d)-14(u)	25(d)-21(u)	25(d)-25(u)	25(d)-25(u)	25(d)-25(u)	25(d)-25(u)	25(d)-25(u)	26(d)-26(u)	26(d)-26(u)	26(d)-26(u)	26(d)-26(u)	
Active T1 links to T2s:	47	47	47	47	47	47	47	47	47	47	47	47	48	48	48	49	53	58	60	62	63	
Waiting Room:	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	out	out	out	out	out	out	out	out	out	out	
	04 Oct	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Report made on 2015-10-25 17:35:02 (UTC)

* = Due to operational errors, the metric has been corrected manually (!=SAB).

"Site Readiness Status" as defined in Site-Readiness Twiki (click me):
- R = READY
- W = WARNING
- NR = NOT-READY
- SD = SCHEDULED-DOWNTIME

"Daily Metric" as boolean AND of all invidual metrics:
- O = OK (All individual metrics above Site Commissioning Thresholds; "n/a" ignored)
- E = ERROR (Some individual metrics below Site Commissioning Thresholds)
- SD = SCHEDULED-DOWNTIME

- INDIVIDUAL METRICS -

"Maintenance": Sites scheduled downtimes
- Up = Site is not declaring Scheduled-downtime
- SD = full-site in SD OR all CMS SE(s) in SD OR all CMS CE(s) in SD
- ~ = Some SE or CE services (not all) Downtime
- UD = full-site in UD OR all CMS SE(s) in UD OR all CMS CE(s) in UD

"HammerCloud":
- = HC success rate is ≥ 90%
- = HC success rate is < 90%
- ~ = Jobs submitted but not finished
- n/a = HC success rate is n/a

"Good Links":
- = at least half of links have 'good' transfers (i.e. with transfer quality > 60%)
- = Otherwise

"SAM Availability":
- = SAM availability is ≥ 90%
- = SAM availability is < 90%

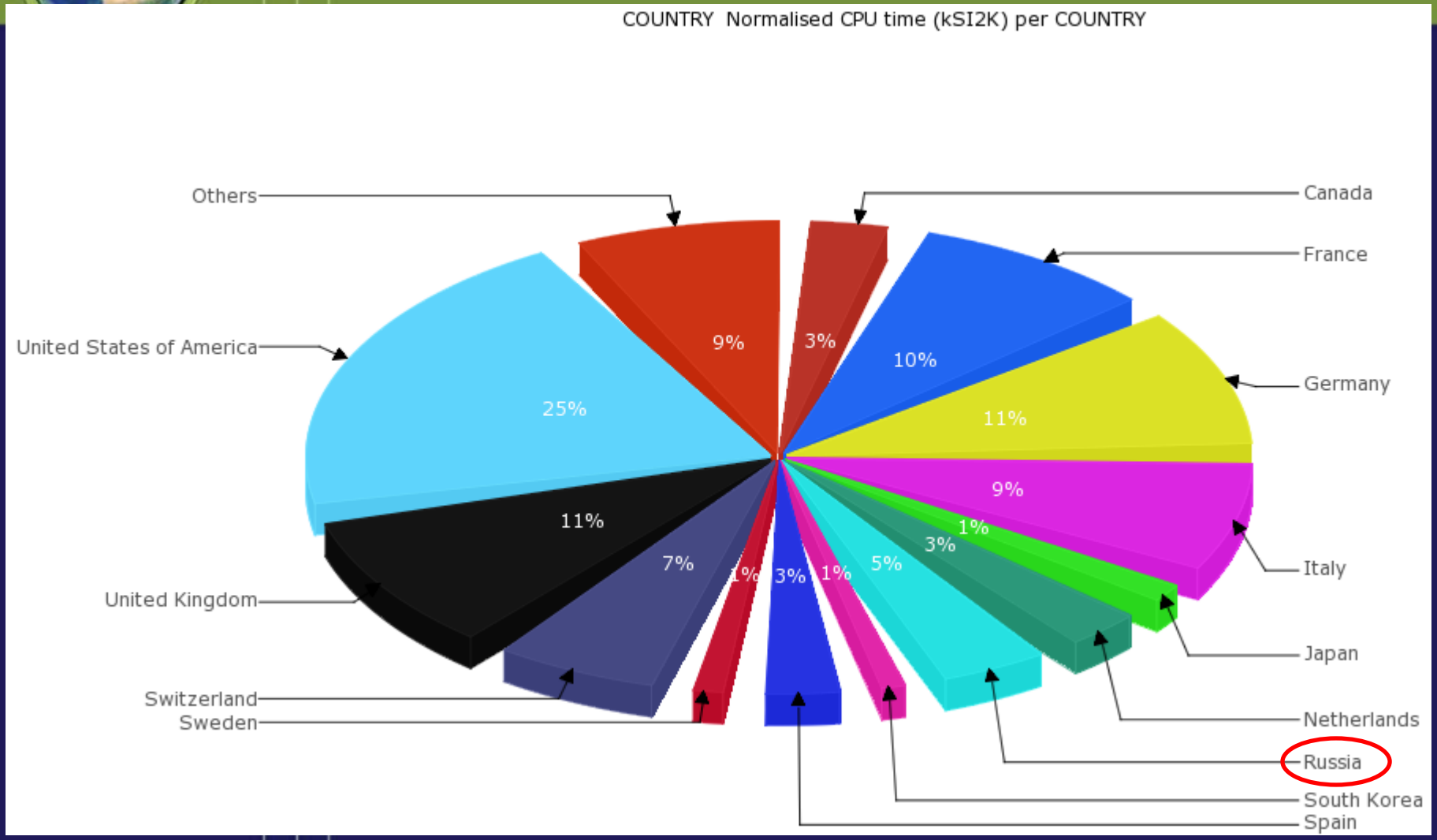
"Active T1 links from T0":
- = Link from T0_CH_CERN is DDT-commissioned
- = Otherwise

"Active T1 links from/to T1s":
- = Site has ≥ 4 DDT-commissioned links from and to, respectively, other T1 sites
- = Otherwise

"Active T1 links to T2s":
- = Site has ≥ 20 DDT-commissioned links to T2 sites
- = Otherwise

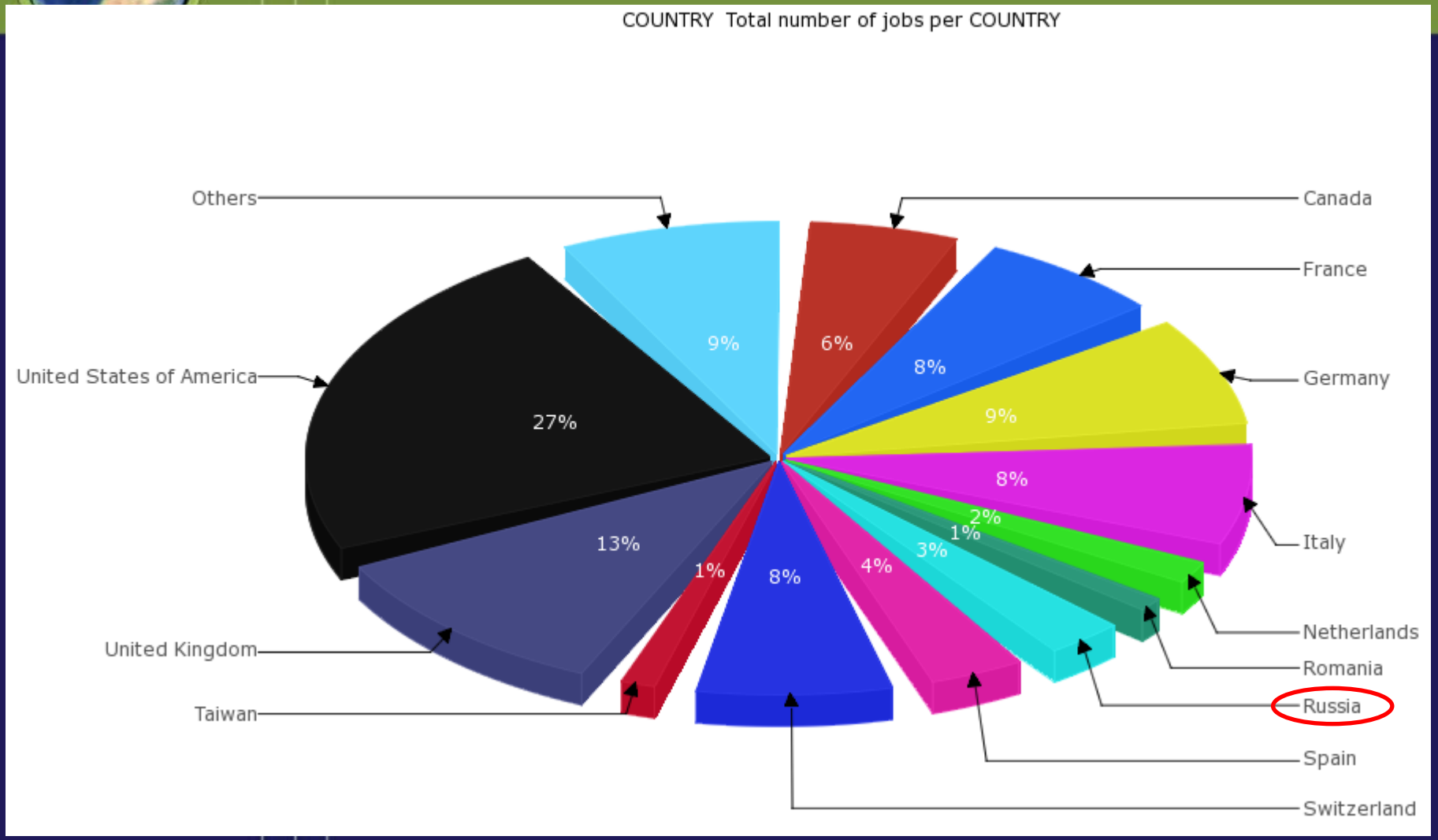
Investment of Russia T1+T2 sites into the total CPU time produced for LHC jobs

November, 2014 - October, 2015



Investment of Russia T1+T2 sites into the total number of LHC jobs run

November, 2014 - October, 2015

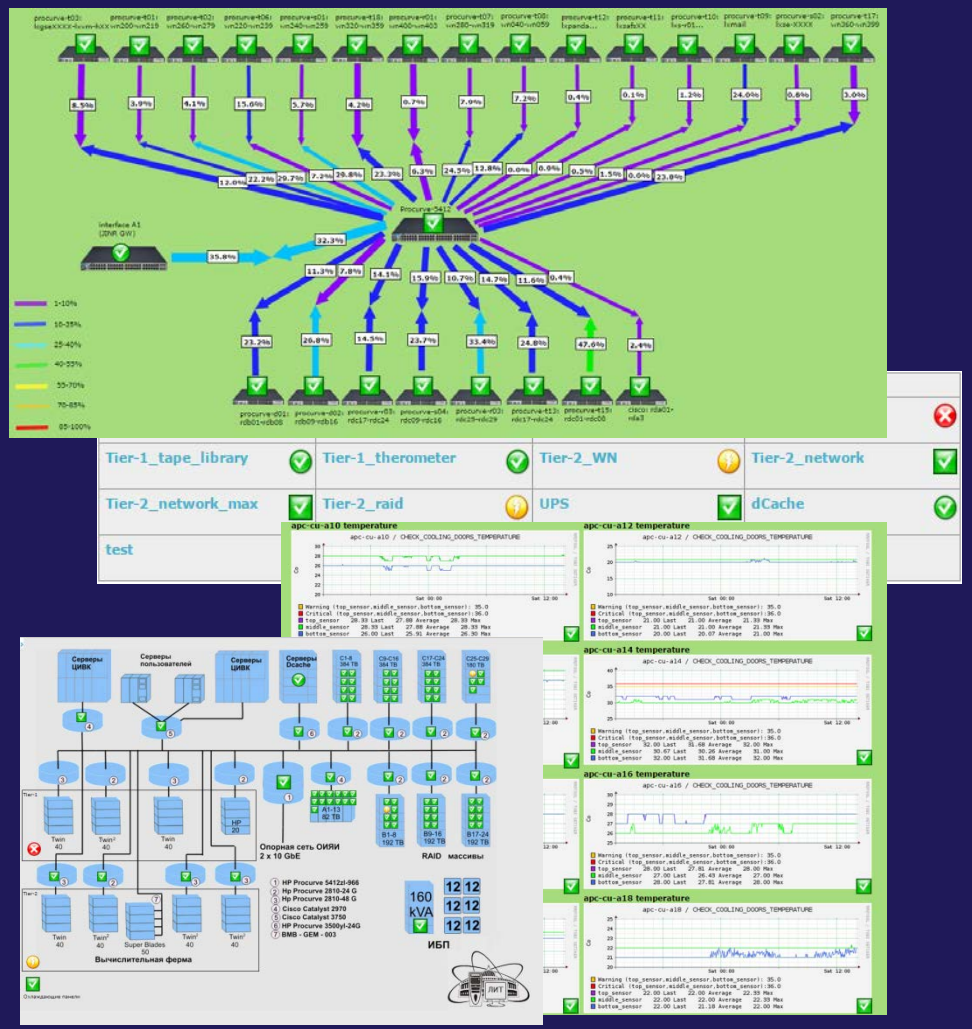


JINR Tier-1 monitoring system

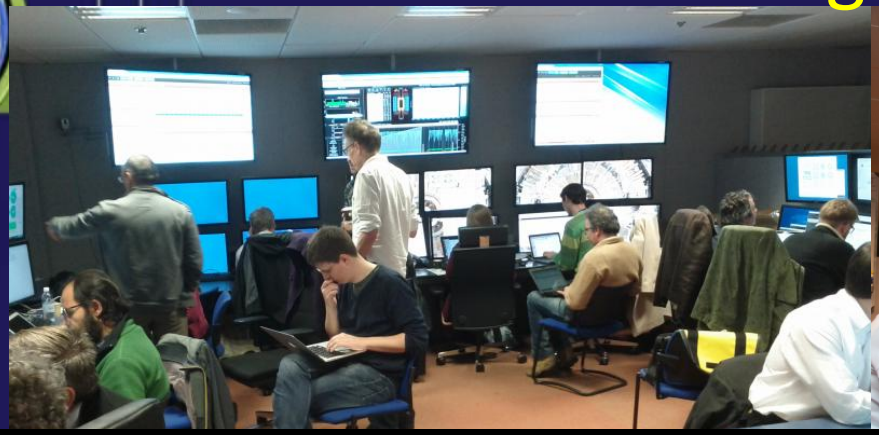
JINR Tier-1 monitoring system provides real-time information about:

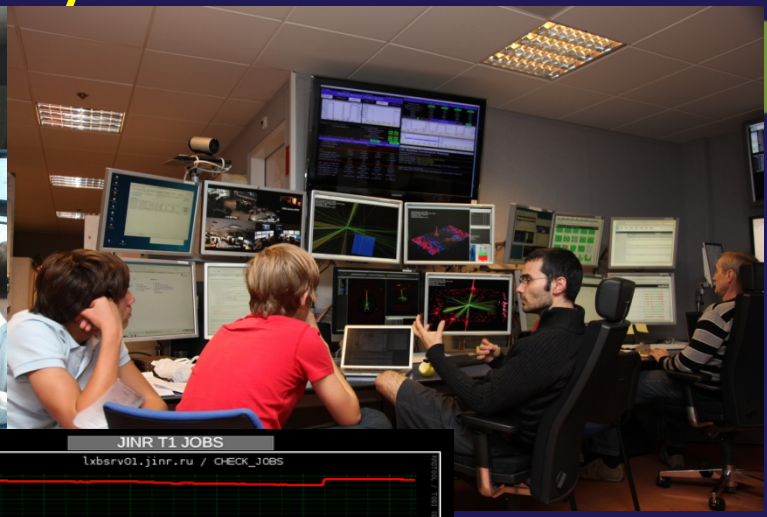
* work nodes;
* disk servers;
* network equipment;
* uninterruptible power supply elements;
* cooling system.

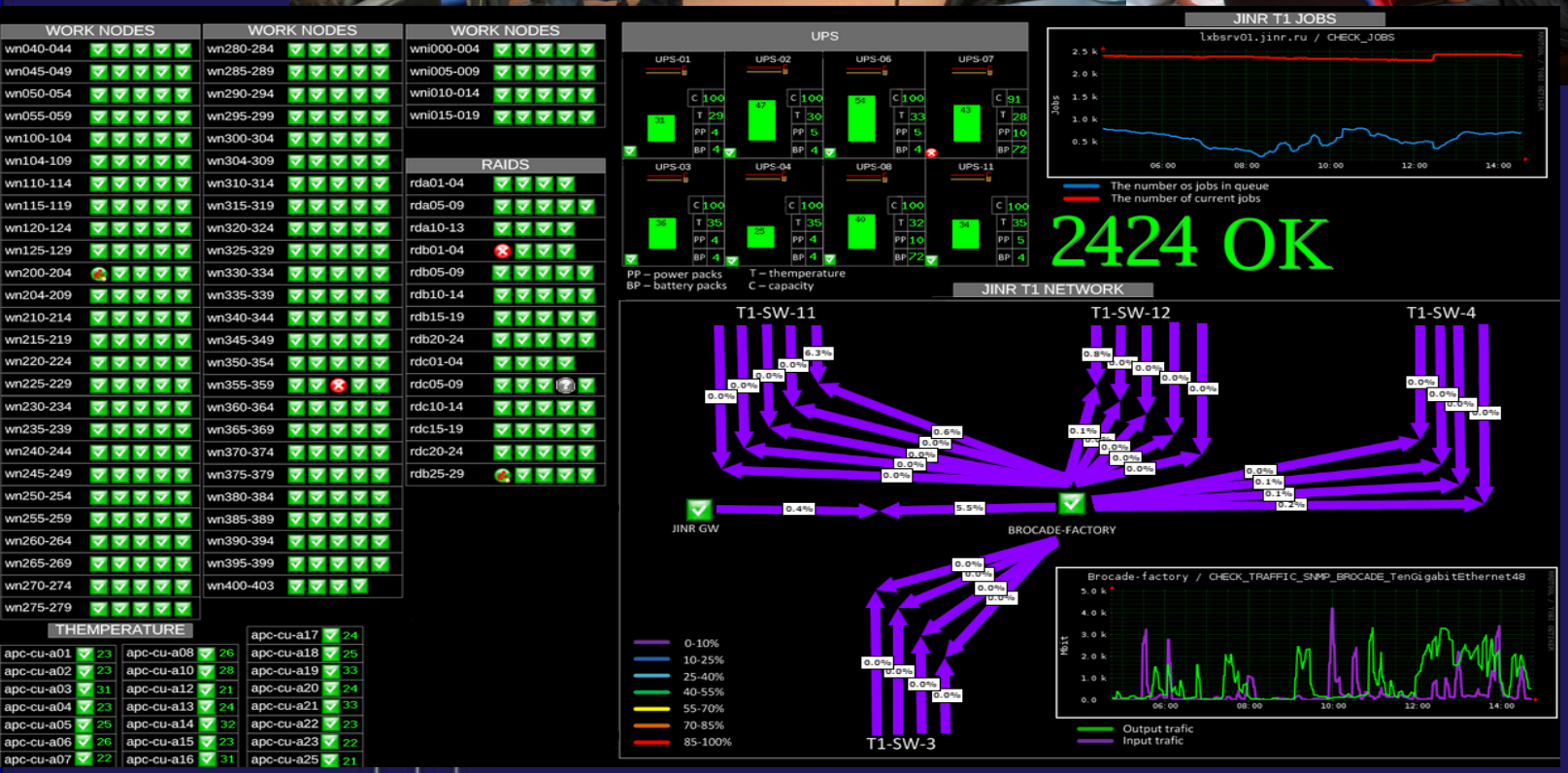
It also can be used for creating network maps and network equipment load maps, for drawing state tables and different plots.



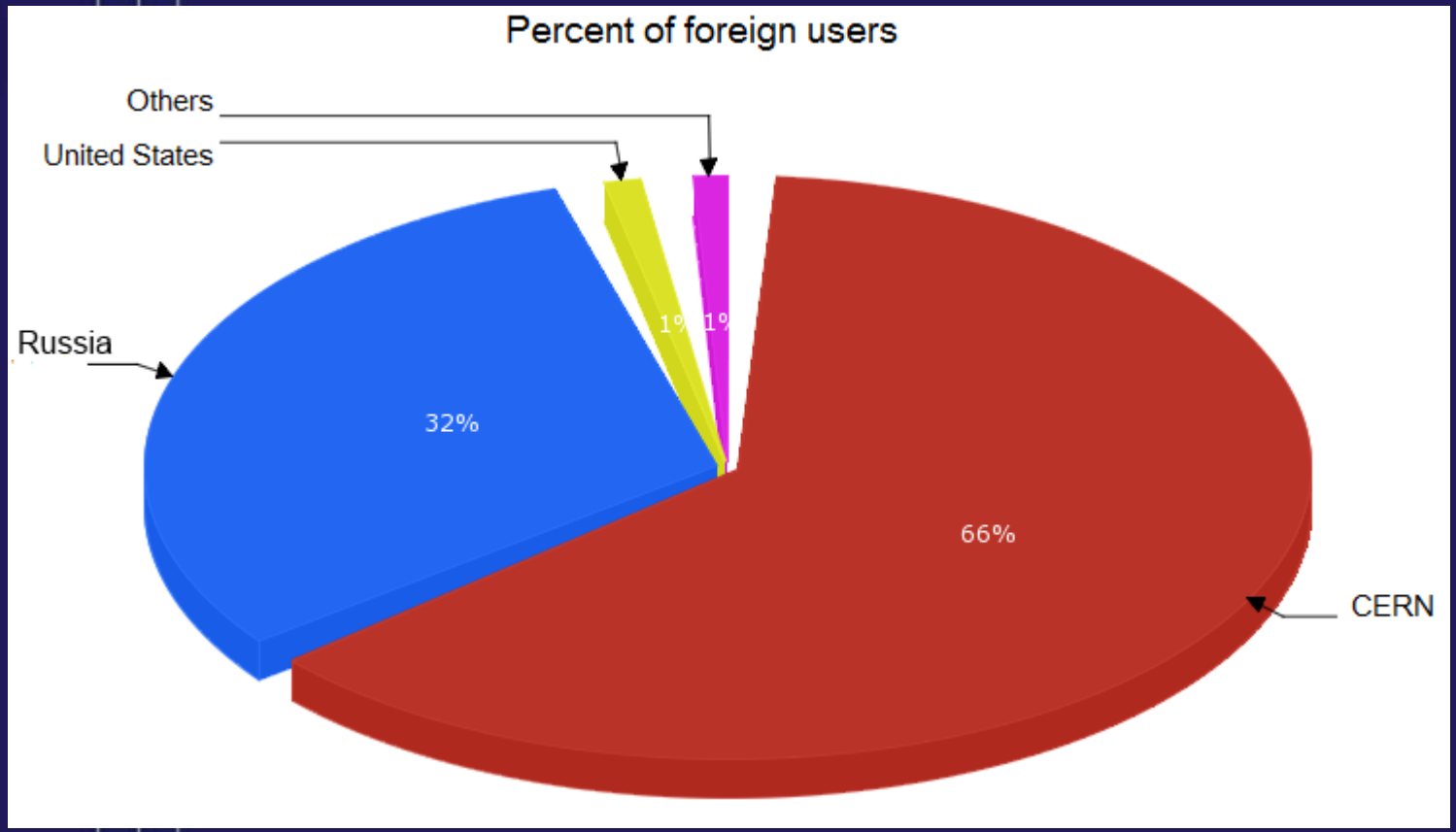
Development perspectives of the JINR Tier-1 monitoring system



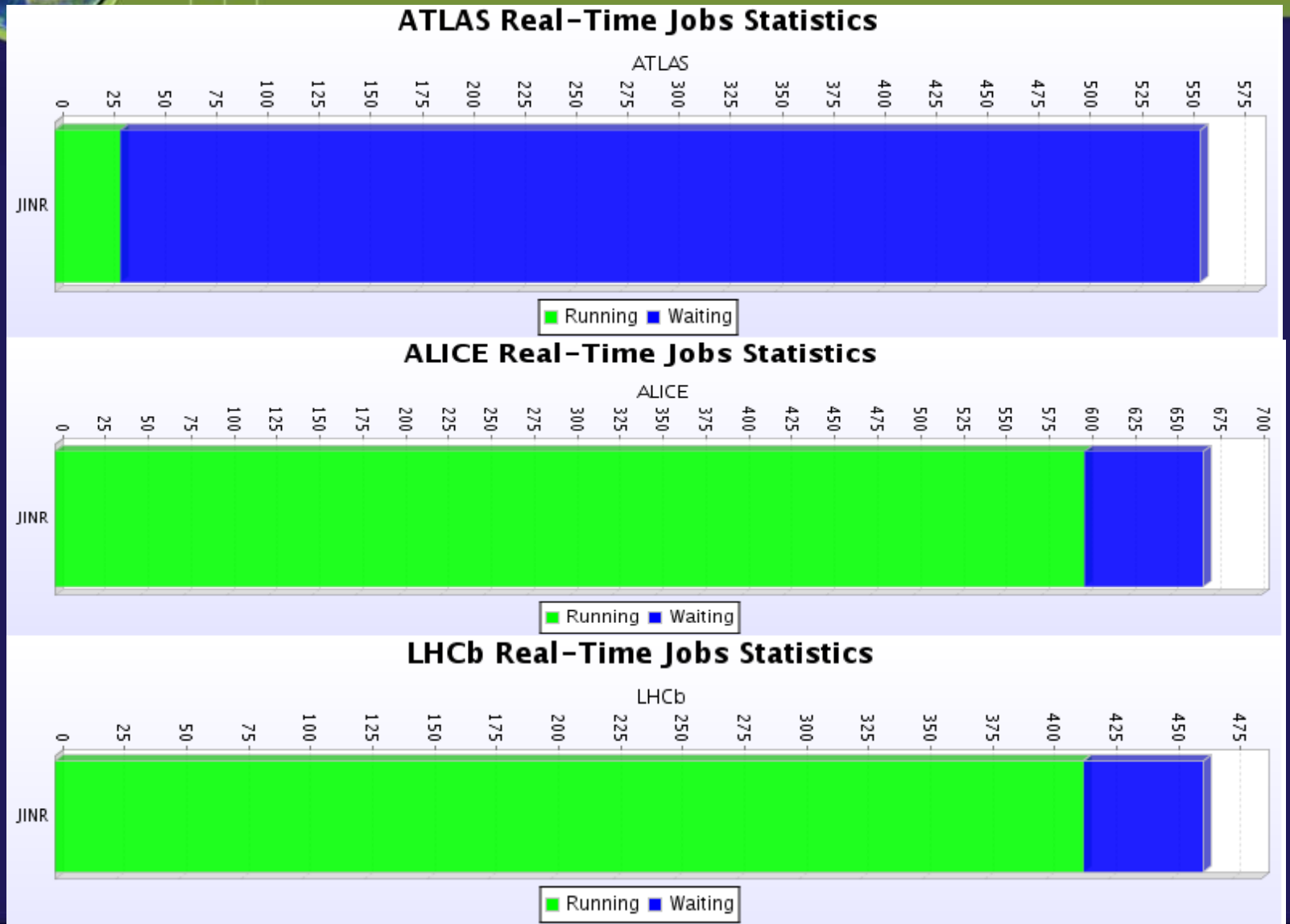




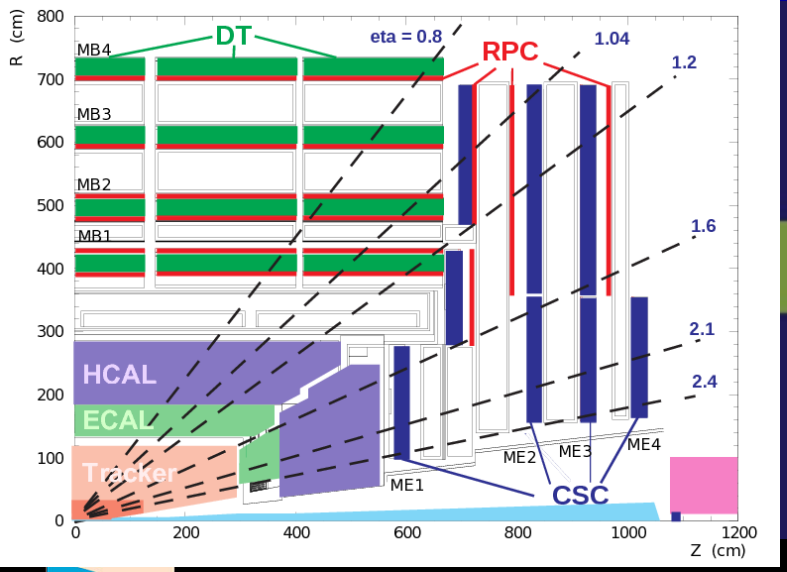
JINR Tier2 CPU resource usage by all VO during November, 2014 – October, 2015

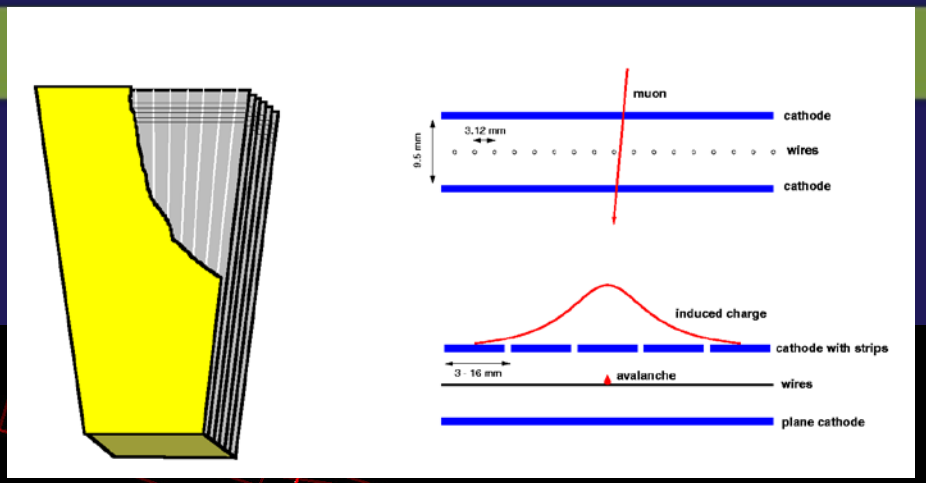


Other major LHC experiments



JINR LIT CMS group
Cathode Strip Chambers (CSC)

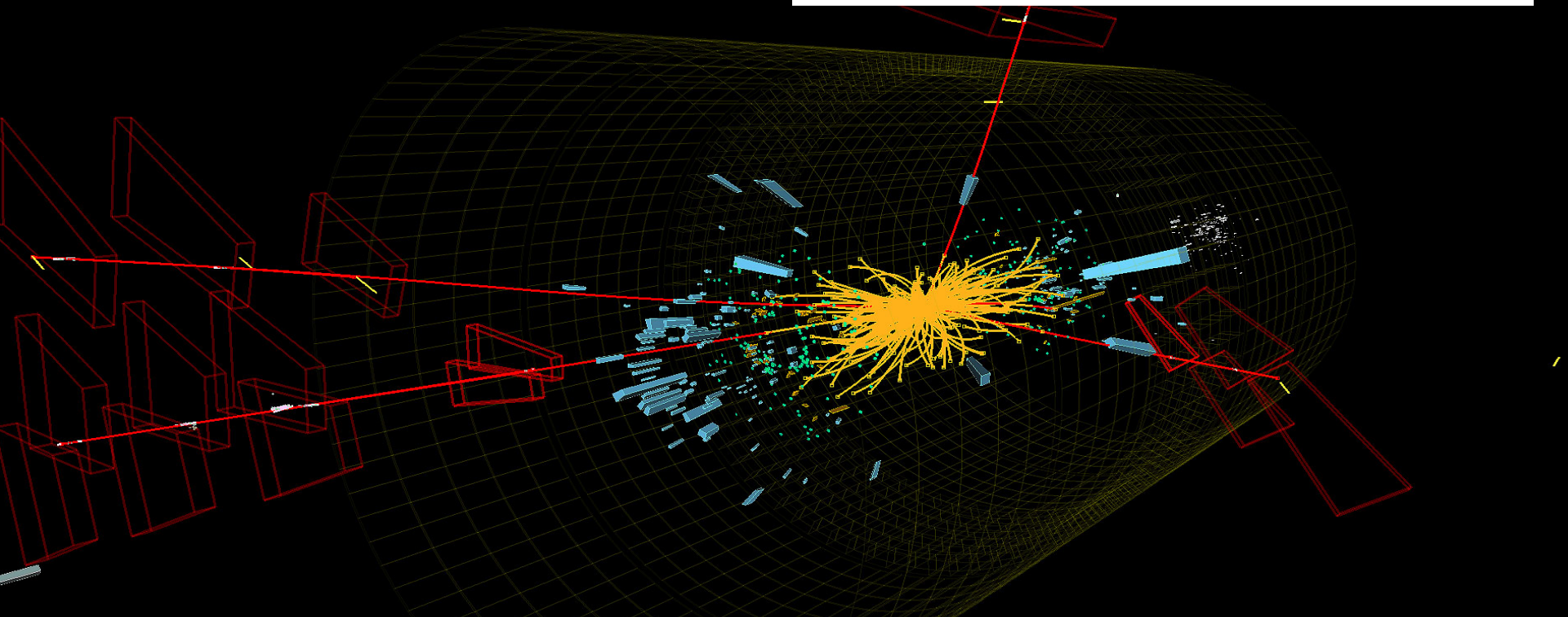




CMS Experiment at the LHC, CERN

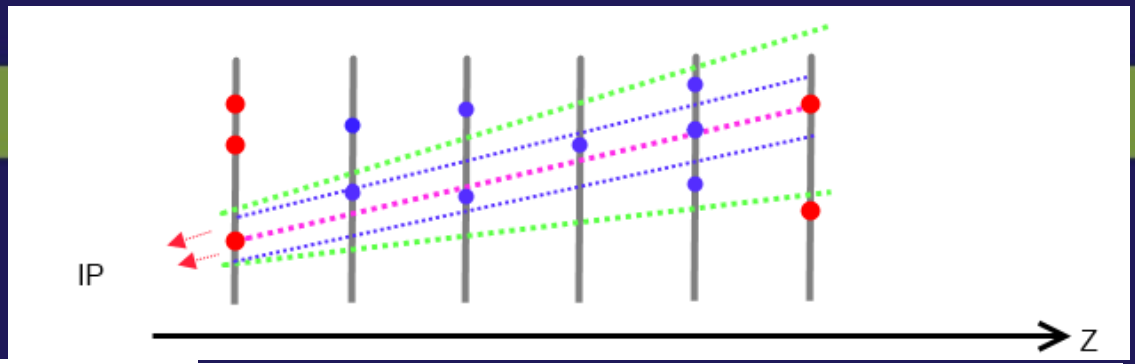
Data recorded: 2011-Jun-05 09:01:21.346043 GMT(04:01:21 CDT)

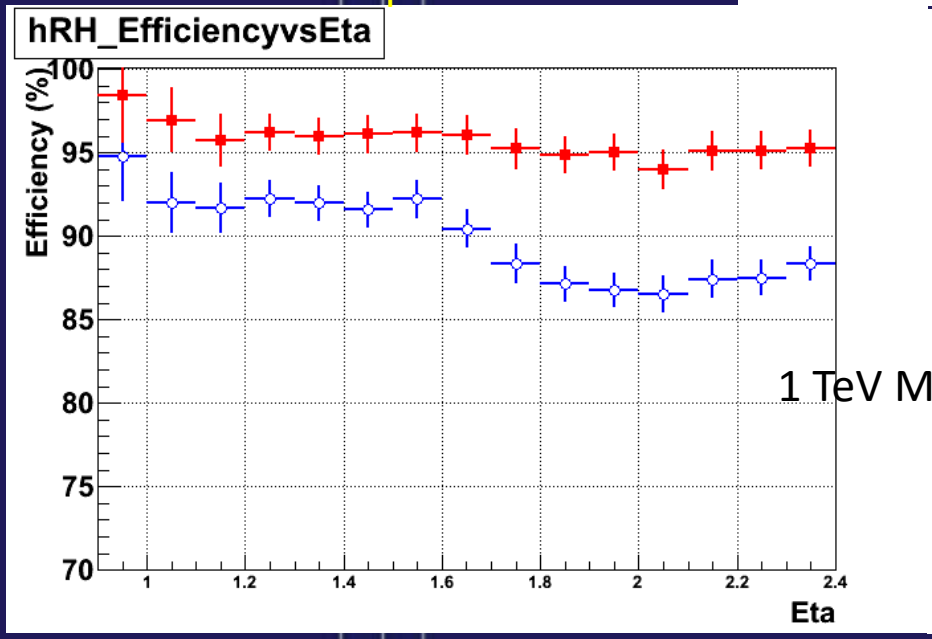
Run / Event: 166512 / 337493970



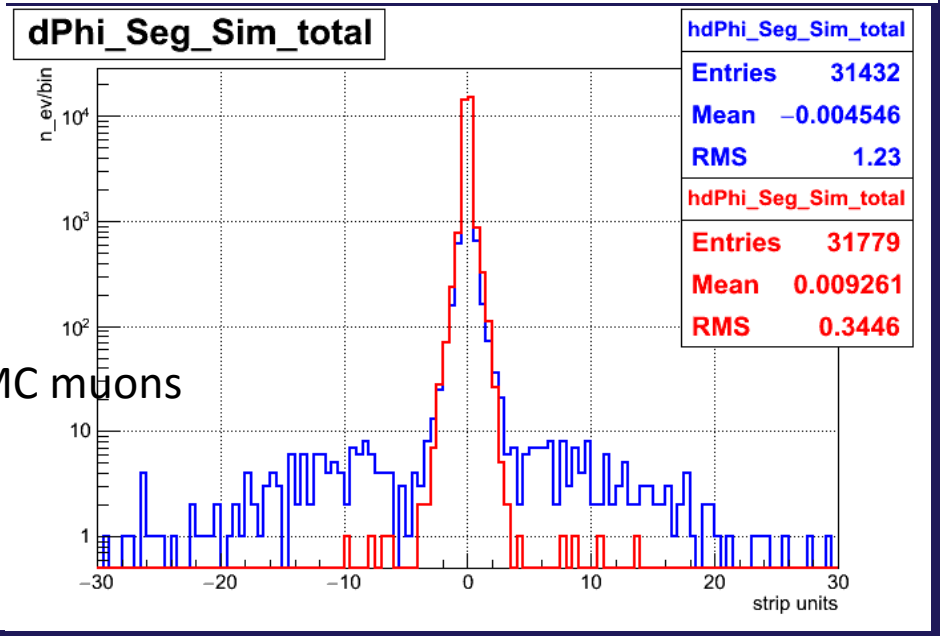
JINR LIT CMS group
Development of a new CSC segment building algorithm

The new CSC segment building algorithm takes into account the interaction point





Reconstruction efficiency vs pseudorapidity

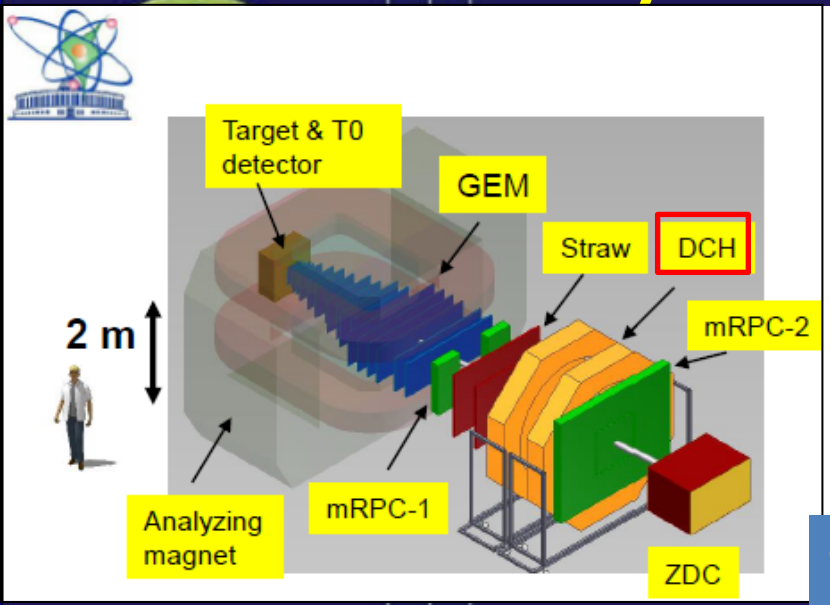


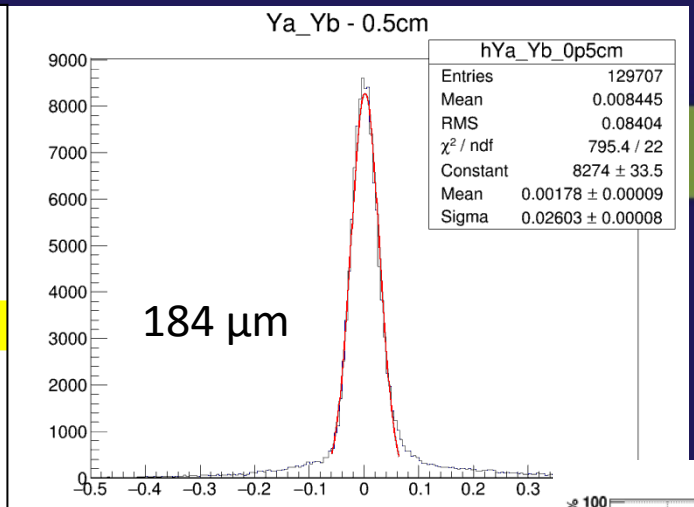
1 TeV MC muons

Difference in angles of the reconstructed and simulated segments (strip_unit ~ 3mrad)

~5000 CRAB jobs
~5TB skimmed datasets

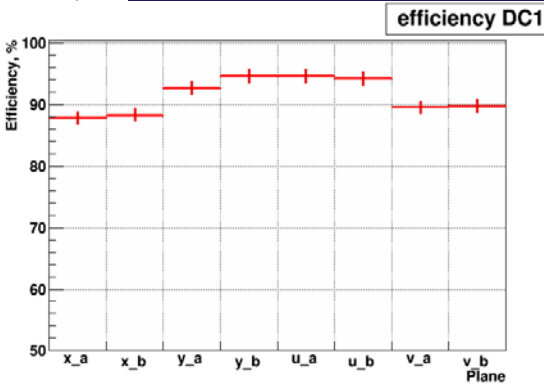
Baryonic Matter at Nuclotron



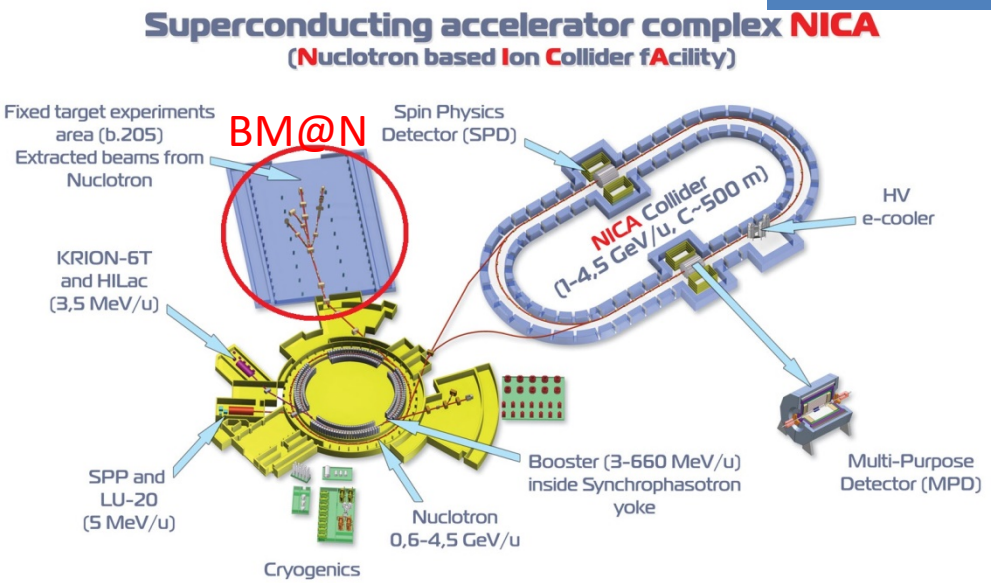


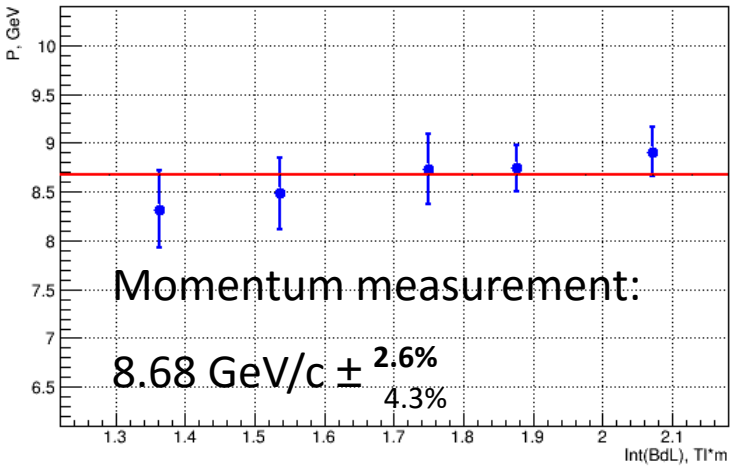
DCH Detector Resolution

DCH Detector Efficiency

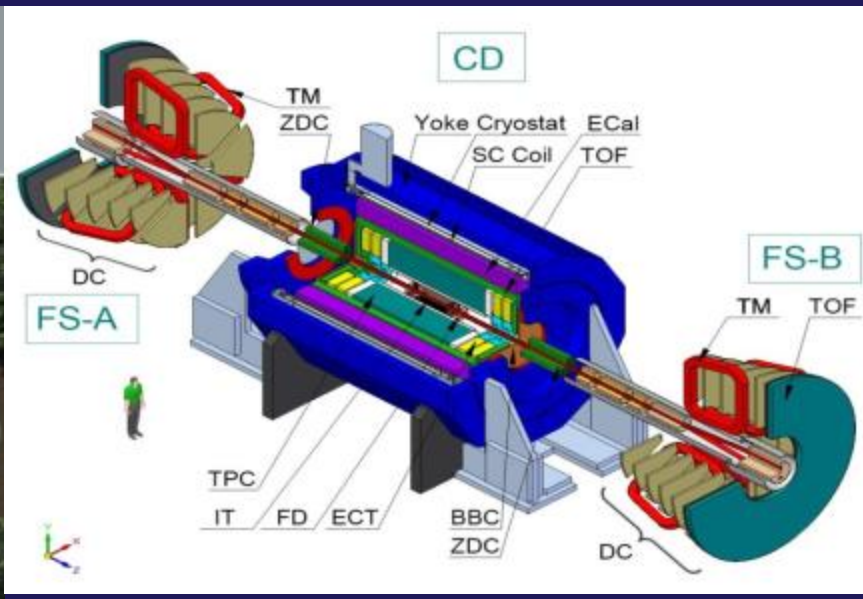


- Fixed target experiment
- Proton and gold beams
- 1-6 GeV per nucleon





NICA Project

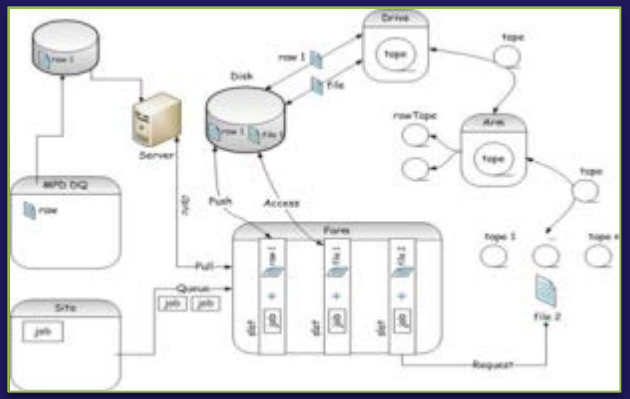


NICA project data amount estimation:

- High frequency of registered events (up to 6 KHz);
- in Au-Au collisions at the expected energies more than 1000 charged particles per event will be produced;
- overall number of events – 19 billions;
- overall amount of data produced per year – 30 PB and 8.4 PB more after analysis.

The model of a distributed computer infrastructure

The model for detailed investigation and estimation:
✓Tape robot,
✓Disk array,
✓CPU Cluster.



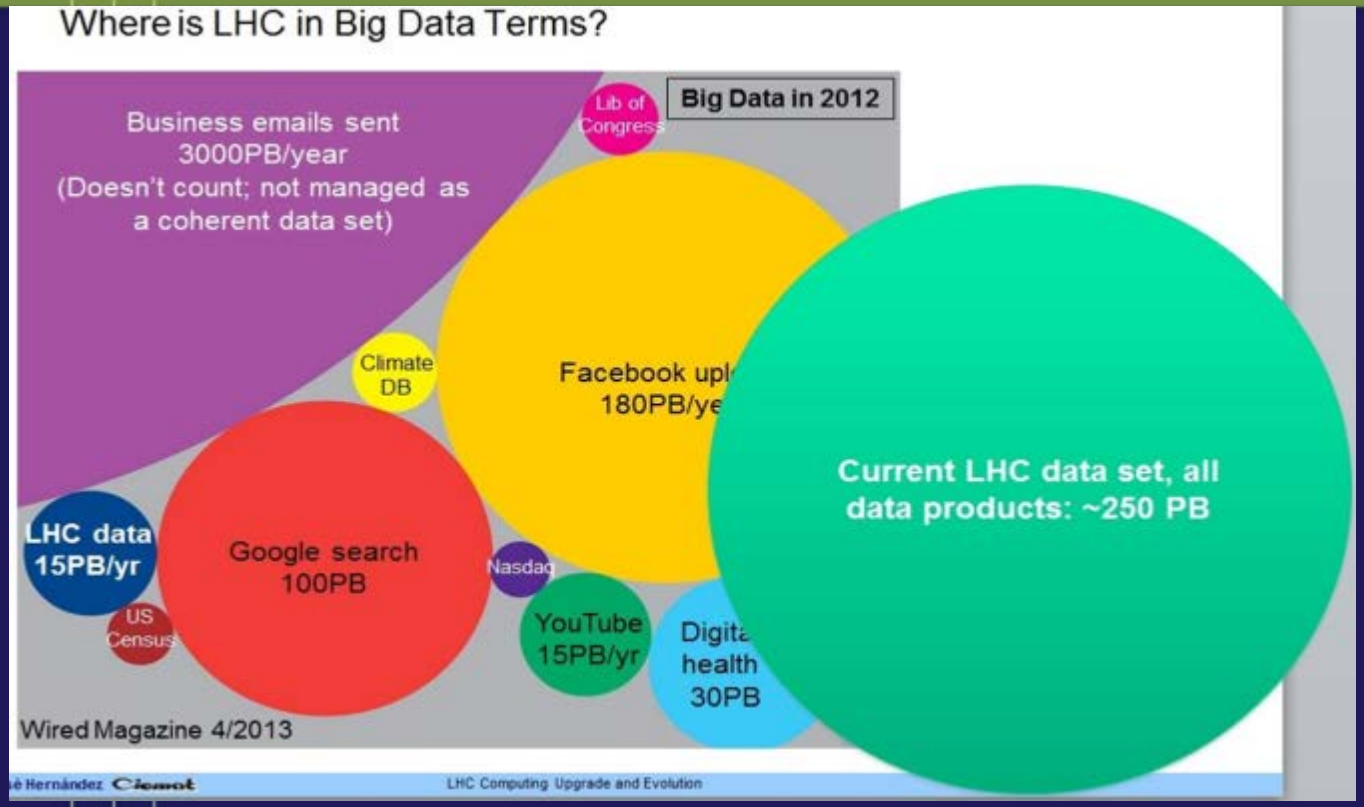
Conclusions

- The concept of grid perfectly fits the LHC project, making possible tremendous amounts of data transfers and job processing.
- The JINR Tier-1 site along with the Russia Tier-2 sites gives the possibility for physicists from JINR, member states to fully participate in processing and analysis of the LHC experimental data.
- An enormous experience in building and maintaining big data processing and storing center was obtained that can be very useful for the development of large scale projects at JINR and other member states.

Thank you for your attention!

Back-up slides

Stepping into Big Data era



The comparison plot of the worldwide processed data shows that the amount of data that comes from the LHC fits the term of Big Data.

The expected amount of data received from LHC should become 2.5 times bigger in Run2, that will require the increase in resources and the optimization of their usage.